The Proliferation and Implications of AI-Generated Child Sexual Abuse Material in the United States: An Expert Analysis for Researchers and Specialized Units

I. Introduction: The Emergence and Escalation of Deepfake CSAM in the United States

A. The Evolving Threat Landscape of Online Child Exploitation

Technological advancements have consistently, albeit often inadvertently, furnished new vectors for child exploitation.[1]
It is also important to consider the nuances between the terms "deepfake" and "AI-generated" in the context of CSAM. "Deepfake," as defined in model legislation, often implies a deliberate attempt to create media that "would falsely appear to a reasonable observer to be an authentic record".14 This suggests an intent to deceive regarding the authenticity of a depicted event or identity. "AI-generated CSAM," on the other hand, is a broader descriptor referring to any CSAM fully or partially created using AI, regardless of whether it is intended to pass as an authentic record of a real event.3 An offender might, for example, create AIG-CSAM that is clearly fantastical or non-realistic in its scenario, yet still illegal and harmful due to its depiction of child sexual abuse. While both fall under the umbrella of CSAM, the specific intent (e.g., deception versus creation of illicit material for other purposes) and the perceived authenticity could influence legal arguments, prosecutorial strategies, and public awareness messaging. Legal and policy frameworks must be precise in addressing whether they are targeting the deceptive nature inherent in some deepfakes, the illicit content itself, or both.

A. The Evolving Threat Landscape of Online Child Exploitation

The digital age has witnessed a continuous adaptation by offenders, leveraging emergent technologies to perpetrate and conceal abuse. In this context, the rapid improvement and increasing accessibility of Artificial Intelligence (AI) models have introduced a new dimension to this threat. While AI offers transformative potential across numerous sectors, its misuse has led to severe harms, among which nonconsensual intimate imagery (NCII) and Child Sexual Abuse Material (CSAM) represent some of the most egregious applications.[3] This evolving landscape necessitates a constant re-evaluation of risks and response strategies by professionals dedicated to child protection.

The capabilities of AI, particularly in image and video synthesis, have become increasingly sophisticated and, critically, more democratized. This democratization of sophisticated abuse tools, where advanced AI models [3] and open-source deepfake software [6] become readily available, signifies a perilous shift. The barrier to entry for creating highly realistic synthetic media has been substantially lowered. Tools like Stable Diffusion and adaptable models such as Checkpoint and LoRA can be downloaded and utilized offline, circumventing content moderation and detection mechanisms that might be present on centralized platforms.[6] Consequently, individuals with even limited technical proficiency can now generate convincing AI-generated CSAM (AIG-CSAM), thereby expanding the potential pool of offenders beyond a small cadre of highly skilled actors and broadening the overall threat landscape.

B. Emergence of Deepfake Technology as a Tool for CSAM Creation and Dissemination

Deepfake technology, which employs AI to create synthetic media where a person's likeness is depicted saying or doing something they never did, was not initially conceived for malicious purposes. However, its potential for misuse was rapidly identified and exploited across various domains, including political disinformation, fraud, and the creation of non-consensual pornography.[2] Its application in the realm of child exploitation marks a particularly disturbing development. This has led to a significant shift from traditional CSAM, which necessitates the direct abuse of a real child for each piece of material, to AIG-CSAM. AIG-CSAM can involve the creation of entirely new, hyper-realistic depictions of non-existent children, the manipulation of existing innocent images of real children, or the alteration of pre-existing CSAM.[2] This distinction is fundamental to understanding the novel challenges that AIG-CSAM poses to law enforcement, victim support services, and legal frameworks.

Furthermore, deepfake technology does not merely introduce a novel category of CSAM; it acts as an amplifier for existing problems within the child protection ecosystem. The sheer volume of digital material that law enforcement must already process is exacerbated by the influx of AIG-CSAM, making it more difficult to identify real children in active abuse scenarios.[3] Survivors of previous abuse face the harrowing prospect of re-victimization, as offenders use AI to generate new abusive material featuring their likenesses, often derived from previously circulated images of their abuse.[3] Sextortion schemes, already a significant concern, are further potentiated by AIG-CSAM, where the synthetic nature of the material does little to mitigate the victim's terror and trauma.[3] Thus, the core issues associated with CSAM are not only replicated but magnified by the capabilities inherent in AI-driven synthesis.

The widespread proliferation of synthetic media, encompassing deepfakes used for diverse purposes—ranging from benign entertainment to malicious activities like political manipulation or adult non-consensual pornography [2]—carries a risk of broader societal desensitization. This potential normalization of manipulated content could inadvertently diminish the perceived severity of AIG-CSAM among certain segments of the public or, more alarmingly, among offenders themselves.[11] As digitally altered realities become more commonplace, the unique abhorrence of CSAM might be diluted if AIG-CSAM is subsumed into a general category of "fake content," thereby complicating efforts to maintain a focused societal condemnation and robust legal response specifically against the sexual exploitation of children in synthetic forms.

C. Purpose and Scope of the Report

This report aims to provide a comprehensive, expert-level analysis of the challenges and implications of deepfake CSAM within the United States. It is intended for an audience of academic researchers, specialized law enforcement personnel, particularly those in Child Abuse, Internet Crimes Against Children (ICAC), and Cybercrime (CALC) units, and other domain experts requiring a nuanced understanding of this evolving threat. The scope encompasses a detailed examination of definitions, prevalence and trends, the U.S. legal and regulatory framework, technological dimensions of creation and detection, operational challenges for investigative units, the profound psycho-societal impacts, and strategic recommendations for mitigation and response.

D. Methodology

The findings and analyses presented herein are derived from a systematic review of peer-reviewed academic journals, official reports from governmental and non-governmental child protection organizations, legal analyses, and technical papers focused on AI, deepfakes, and CSAM.[3] Particular attention has been given to data and reports from U.S.-based entities and those with direct relevance to the U.S. context.

II. Defining the Terrain: Understanding CSAM, Deepfakes, and AI-Generated CSAM (AIG-CSAM)

A. Child Sexual Abuse Material (CSAM): Established Definitions and Legal Context

Child Sexual Abuse Material (CSAM) is broadly defined as sexually explicit content involving a child.[3] The U.S. Department of Justice emphasizes that underlying every sexually explicit image or video of a child is an act of abuse, rape, molestation, and/or exploitation, making the material itself a permanent record of the child's victimization.[1] While the term "child pornography" persists in some federal statutes, "CSAM" is the preferred terminology among professionals as it more accurately reflects the inherent abuse depicted and the resultant trauma to child victims.[1]

Crucially for the context of this report, U.S. federal law has, for some time, included provisions that address synthetic imagery. The legal definition of CSAM encompasses "computer-generated images indistinguishable from an actual minor, and images created, adapted, or modified, but appear to depict an identifiable, actual minor".[5] This pre-existing legal language indicates that certain forms of non-photographic or manipulated CSAM were already contemplated within the legislative framework, providing a foundational basis for addressing some types of AIG-CSAM.

B. Deepfake Technology: Mechanisms and Capabilities

Deepfake technology refers to "any visual media created, altered, or otherwise manipulated in a manner that would falsely appear to a reasonable observer to be an authentic record of the individual's actual speech, conduct, or likeness".[14] These synthetic media are typically generated using advanced AI and machine learning (AI/ML) techniques, most notably Generative Adversarial Networks (GANs) and, more recently, diffusion models.[15] These models are trained on large datasets of images or videos to learn and replicate patterns, enabling them to produce highly realistic outputs. While deepfakes have found applications in various fields, their capacity for malicious use—including political disinformation, financial fraud, and the creation of non-consensual pornography involving adults—underscores the dual-use nature of this potent technology.[8]

C. AI-Generated CSAM (AIG-CSAM): Specific Characteristics and Typologies

AI-Generated CSAM (AIG-CSAM) is defined as CSAM that is fully or partially created using artificial intelligence.[3] This category is not monolithic and encompasses several distinct typologies:

1. **Fully Synthetic Depictions:** Images or videos that portray entirely fictional, AI-generated child-like figures engaging in sexually explicit conduct.[2] These may not be based on any specific real child.
2. **Manipulation of Existing Images of Real Children:** This involves taking innocent photographs or videos of real children, often sourced from social media, and using AI to alter them into sexually explicit material. Common techniques include "nudifying" (digitally removing clothing) or face-swapping a child's face onto pre-existing pornographic content.[3]
3. **Alteration of Existing CSAM:** AI can be used to modify pre-existing CSAM, for instance, to obscure the identity of a victim or perpetrator, or to create new abusive scenarios featuring previously victimized children.[3]
4. **AI-Generated Textual CSAM:** Generative AI can also produce text-based CSAM, such as scripts for grooming children, or chatbots designed to simulate sexually explicit conversations with minors.[17]

The legal and policy implications of AIG-CSAM are significantly shaped by the "indistinguishable" threshold present in federal law.[5] As AI technology advances, its capacity to create AIG-CSAM that is virtually indistinguishable from photographic depictions of real children continues to improve.[3] This technological progression means that an increasing volume of AIG-CSAM will likely meet the existing legal definition, potentially broadening the applicability of current statutes without necessitating new legislation for these specific instances. However, this very same hyper-realism concurrently makes the forensic task of differentiating authentic CSAM from AIG-CSAM, or identifying AI manipulations within genuine CSAM, exceptionally challenging for law enforcement and forensic analysts.[19] This creates a critical tension: while the legal net may inherently widen due to technological advancement, the practical ability to enforce those laws through accurate identification and prosecution becomes more complex.

D. The Critical Misconception: Perceived vs. Actual Harm of AIG-CSAM

A pervasive and dangerous misconception is that AIG-CSAM is somehow less harmful than traditional, non-synthetic CSAM.[3] Experts across the fields of child protection, law enforcement, and academia unequivocally reject this assertion.[3] The harms associated with AIG-CSAM are profound and, in some respects, compounded:

- **Strain on Resources:** The proliferation of AIG-CSAM further taxes an already overburdened child safety ecosystem. Law enforcement agencies face increased difficulties in sifting through vast quantities of material, determining if a real child is depicted, whether AI has been used to conceal an identity, or if the depicted sexual violence is authentic.[3]
- **Re-victimization:** Malicious actors are actively fine-tuning AI models to generate new CSAM featuring survivors of past abuse, thereby perpetuating and intensifying their trauma.[3]
- **Sextortion:** AIG-CSAM is increasingly used in sextortion schemes. For the child victim being blackmailed with synthetically generated explicit images (often created from their innocent social media photos), the "reality" or "fakeness" of the image is irrelevant to the profound feelings of violation, fear, isolation, and the worry that others will not believe the content is fabricated.[3]
- **Psychological Impact:** Even when AIG-CSAM depicts a child synthetically, or uses a real child's likeness without their involvement in an actual abusive act during the image's creation, the psychological toll is severe. Victims report experiencing humiliation, shame, anger, a sense of violation, and self-blame. This can lead to lasting emotional distress, withdrawal from social and academic life, and difficulties in forming trusting relationships.[11]

The creation and dissemination of AIG-CSAM also challenge and expand traditional notions of victimhood. While the direct physical abuse that characterizes the production of traditional CSAM might not occur in the creation of *purely synthetic* AIG-CSAM (i.e., depicting entirely fictional children), the use of a real child's likeness, even if sourced from an innocent photograph, constitutes a distinct form of abuse and inflicts significant psychological and reputational harm.[3] The trauma experienced by a child whose face is synthetically placed into an abusive scenario is palpable and undeniable. Furthermore, even AIG-CSAM that depicts entirely fictional child-like entities contributes to the broader CSAM ecosystem by fueling demand, normalizing the sexualization of children, and potentially desensitizing viewers and creators to the inherent criminality and harm of such material.[1] This broader impact means that the "victim" category extends beyond those directly depicted, implicating all children who are made more vulnerable by the perpetuation of a market for child exploitation.

It is also important to consider the nuances between the terms "deepfake" and "AI-generated" in the context of CSAM. "Deepfake," as defined in model legislation, often implies a deliberate attempt to create media that "would falsely appear to a reasonable observer to be an authentic record".[14] This suggests an intent to deceive regarding the authenticity of a depicted event or identity. "AI-generated CSAM," on the other hand, is a broader descriptor referring to any CSAM fully or partially created using AI, regardless of whether it is intended to pass as an authentic record of a real event.[3] An offender might, for example, create AIG-CSAM that is clearly fantastical or non-realistic in its scenario, yet still illegal and harmful due to its depiction of child sexual abuse. While both fall under the umbrella of CSAM, the specific intent (e.g., deception versus creation of illicit material for other purposes) and the perceived authenticity could influence legal arguments, prosecutorial strategies, and public awareness messaging. Legal and policy frameworks must be precise in addressing whether they are targeting the deceptive nature inherent in some deepfakes, the illicit content itself, or both.

III. Prevalence, Trends, and Statistical Landscape of Deepfake CSAM in the U.S.

A. NCMEC CyberTipline Reporting Trends

The National Center for Missing and Exploited Children (NCMEC) serves as the primary U.S. clearinghouse for reports of suspected online child exploitation. Its CyberTipline data provides critical insights into the scale and evolution of CSAM.
In 2023, NCMEC received over 100 million reports of suspected CSAM.3 This figure, largely predating the significant surge in AIG-CSAM, illustrates the immense baseline workload faced by child protection entities.

For 2024, NCMEC reported receiving 20.5 million reports, which, when adjusted to reflect distinct reported incidents, amounted to 29.2 million separate incidents of child sexual exploitation submitted to the CyberTipline.17 This represented a decrease from the 36.2 million reports (incidents) in 2023. NCMEC attributes this decline partly to a new "bundling" feature implemented in 2024, allowing large online platforms to consolidate related reports of widespread incidents (e.g., viral meme content), thereby reducing redundant submissions.22 However, NCMEC also expressed concern that the decrease might reflect underreporting by some Electronic Service Providers (ESPs) and the growing challenge posed by end-to-end encryption, which can limit platforms' ability to detect and report illicit content, despite legislative efforts like the REPORT Act aimed at increasing such reports.22
Of particular relevance is the dramatic increase in reports involving Generative AI (GAI).

- The NCMEC CyberTipline saw a **1,325% increase** in reports involving Generative AI in 2024, escalating from approximately 4,700 such reports in 2023 to **67,000 reports in 2024**.[17]
- Earlier NCMEC communications (March 2024) specified that the 4,700 reports received in 2023 were related to "Child Sexual Abuse Material (CSAM) or sexually exploitative content that involved GAI technology (GAI CSAM)".[18]
- Separately, the organization Enough Abuse® reported that, as of April 2025, NCMEC had received over 7,000 reports of CSAM specifically involving generative AI technology over the preceding two years.[10]

The phrasing "reports involving Generative AI" for the 67,000 figure in 2024 is broader than "GAI CSAM" and may encompass various applications of GAI in child exploitation, such as AI-powered grooming bots or the generation of instructional materials for abuse, in addition to AIG-CSAM itself.[17] This distinction is important for a precise understanding of the statistics.

Table 1: NCMEC CyberTipline Key Statistics (2023-2024)

Reporting Category	2023 Reports (Incidents)	2024 Reports (Incidents)	2024 % Change from 2023 (Incidents)	Notes
Total Suspected CSAM/Exploitation Reports (Incidents)	36.2 million	29.2 million	Approx. -19.3%	2024 figures reflect bundling; NCMEC expresses concern over potential underreporting by some ESPs.
Reports Involving Generative AI (GAI)	~4,700	67,000	+1,325%	"Involving GAI" is broad; includes GAI CSAM and other GAI uses in exploitation.
Online Enticement Reports	(2023 data not specified)	>546,000	+192% (compared to 2023 reports)	NCMEC anticipates continued growth.
Child Sex Trafficking Reports	(2023 data not specified)	26,823	+55% (compared to 2023 reports)	Increase potentially an early result of the REPORT Act.
Reports with Nexus to Violent Online Groups	(2023 data not specified)	>1,300	+200% (over 2023)	Majority (69%) from parents/caregivers after child's self-harm/suicide attempt.

Sources:.[3] *Note: Specific incident counts for 2023 for enticement, trafficking, and violent groups were not available in the provided snippets, only percentage increases for 2024 compared to 2023.*

This table provides a crucial quantitative snapshot. For experts, the dramatic rise in AI-related reports, juxtaposed with overall CSAM figures and contextual data like online enticement, powerfully illustrates AI's rapid emergence as a significant factor in child exploitation. The distinction between the 67,000 "GAI involvement" reports and more specific "GAI CSAM" figures (like the earlier 4,700 for 2023 or the 7,000 over two years) highlights both explosive growth and the ongoing need for precise definitions and disaggregated data in reporting to accurately track specific AIG-CSAM typologies. The impact of report bundling on overall CSAM figures is also vital context, preventing misinterpretation of raw decreases as a reduction in actual offenses.

B. Other Reporting and Research Findings on Prevalence

Beyond NCMEC, other organizations and researchers have shed light on the prevalence of AIG-CSAM:

- The **Internet Watch Foundation (IWF)**, a UK-based organization, discovered over 20,000 AI-generated child abuse images on a single dark web forum within a one-month period in late 2023/early 2024.[20] This finding, while not U.S.-specific, indicates the high-volume production capabilities and distribution occurring on certain illicit platforms, likely accessible globally.
- A 2024 survey by **Thorn**, a U.S. non-profit, found that 1 in 10 minors (11%) in the United States, aged 9 to 17, reported knowing peers who had used AI to create sexually explicit images of other minors.[9] WeProtect Global Alliance also cited a Thorn report indicating that 1 in 10 children are aware of peers using generative AI to create non-consensual intimate images of others.[24] This points to peer-on-peer creation as an alarming and significant emerging trend.
- Academic research corroborates that AI technologies are increasingly employed for CSAM fabrication, including "nudifying" innocent pictures and manipulating images to depict known or unknown children in sexually abusive contexts.[9]

The official statistics from NCMEC, while alarming, likely represent only a fraction of the true volume of AIG-CSAM. The ease of offline generation using open-source tools [7] means a significant amount of material can be created without leaving an immediate digital footprint on monitored platforms. Coupled with the potential for underreporting by some service providers due to encryption or other factors [22], a substantial "dark figure" of AIG-CSAM production and circulation is highly probable. The IWF's finding of over 20,000 images on a single dark web forum in one month [20] offers a stark glimpse into this largely unmonitored ecosystem, suggesting the scale of unreported AIG-CSAM could be vast.

C. Key Trends in AIG-CSAM

Several key trends characterize the AIG-CSAM phenomenon:

- **Increasing Realism and Sophistication:** AIG-CSAM is rapidly evolving to become visually indistinguishable from authentic CSAM, even for trained forensic analysts.[6] This hyper-realism significantly complicates detection, investigation, and prosecution.
- **Offline Generation:** The ability for perpetrators to download open-source AI models and generate CSAM offline, on local devices, allows them to evade online detection mechanisms employed by platforms and law enforcement until the material is distributed.[7]
- **Use of Real Children's Images as Source Material:** A prevalent method involves taking innocent photographs of children, often sourced from their social media profiles or those of their families, and using these as base images for creating AIG-CSAM.[2]
- **Re-victimization of Survivors:** Existing CSAM is being used to train AI models, or original images of abuse are being altered by AI, to create new, synthetic depictions of children who have already been victimized.[3]
- **Emergence of AI-Generated CSAM Videos:** Beyond static images, realistic deepfake videos depicting child sexual abuse are now circulating, representing a further escalation in the sophistication of AIG-CSAM.[20]
- **Commercialization and Dissemination Tactics:** AIG-CSAM is presenting new avenues for perpetrators to profit from exploitation. This includes the sale of AIG-CSAM on commercial sites or through subscription

models on content-sharing platforms, and the distribution of guides on how to generate such material.[6] Dark web forums remain key locations for sharing intelligence and techniques related to AIG-CSAM production.[7]

- **Peer-on-Peer Creation:** The data indicating that minors are using AI to create non-consensual explicit images of their peers [9] signals a disturbing shift in offender demographics and motivations. This trend suggests that the creation of AIG-CSAM is not confined to adult predators. Motivations among youth may differ, potentially including bullying, social status, curiosity, or a profound lack of understanding of the severity and illegality of their actions. This blurring of lines between victim and offender in some youth contexts necessitates tailored prevention, education, and intervention strategies distinct from those targeting adult offenders.

IV. The U.S. Legal and Regulatory Framework: Addressing Deepfake CSAM

The legal response to AIG-CSAM in the United States is evolving, characterized by interpretations of existing federal statutes, a rapid proliferation of state-level legislation, and ongoing constitutional debates.

A. Federal Legislative Landscape

Existing federal CSAM laws, primarily codified under 18 U.S.C. § 2251 et seq., broadly prohibit the production, distribution, receipt, and possession of CSAM.[1] As noted earlier, the federal definition of CSAM already includes "computer-generated images indistinguishable from an actual minor, and images created, adapted, or modified, but appear to depict an identifiable, actual minor".[5] In 2024, the Federal Bureau of Investigation (FBI) publicly affirmed that AIG-CSAM falls under this definition and is considered CSAM.[5]

Despite this affirmation, a significant legal nuance persists: as of early 2025, there have been no reported instances of a U.S. federal case being brought *solely* based on AIG-CSAM that does *not* depict an identifiable, actual child and was *not* trained using imagery of real children engaged in sexual abuse.[5] This highlights an area of untested application of federal law or potential legal uncertainty regarding purely synthetic AIG-CSAM that lacks a direct link to a specific, real-world child victim.

Several pieces of federal legislation have been enacted or introduced to address deepfakes and AI-generated content more broadly, which could have implications for AIG-CSAM:

- **The TAKE IT DOWN Act (S. 146 / H.R. 633)** was signed into law on **May 19, 2025**. It aims to prohibit the nonconsensual disclosure of AI-generated intimate imagery (NCII, which can include deepfakes) and mandates that online platforms remove such content upon notification. While the Act represents a significant legislative step, it has also faced criticism, notably from the Electronic Frontier Foundation (EFF), which argues that its broad definitions, rapid takedown mandates (within 48 hours), and lack of robust safeguards may pose risks to free expression, user privacy (particularly with end-to-end encrypted services), and due process. This ongoing debate underscores the tension between protecting victims from harmful synthetic media and safeguarding fundamental digital rights.

B. State-Level Legislative Responses

In contrast to the more measured federal pace, U.S. states have responded with notable urgency to the threat of AIG-CSAM.

- As of April 2025, **38 states** have enacted laws that explicitly or implicitly criminalize AI-generated or computer-edited CSAM.[10] Significantly, more than half of these laws were passed in 2024 alone, reflecting escalating concern among state legislators.
- These state laws exhibit considerable **variation** [10]:
  - Some states, such as **California** (e.g., AB 1831, SB 1381, effective January 1, 2025), have amended their existing child pornography statutes to explicitly include matter that is "digitally altered or artificial-intelligence-generated" depicting a minor in sexual conduct.[10]
  - Other states utilize broader language, criminalizing material "morphed or produced by electronic or other means," which can be interpreted to cover AIG-CSAM without specifically naming AI.[10]
  - A common approach is to criminalize AIG-CSAM when an image of an **actual, identifiable child** is used as the basis for the synthetic depiction.[10]
  - However, some states have gone further, defining the prohibited material as any image that **"appears to be a minor"** under 18 involved in sexual activity, irrespective of whether a real child is depicted or can be identified.[10] This "appears to be" standard directly engages with the constitutional questions raised by *Ashcroft v. Free Speech Coalition*.
  - For example, **Alabama's Child Protection Act of 2024 (HB 168)** expands the definition of CSAM to include "virtually indistinguishable depictions" created, altered, or produced by digital, computer-generated, or other means.[28]
- Model legislation, such as the **"Stop Deepfake CSAM Act"** developed by the American Legislative Exchange Council (ALEC) in December 2023, likely influences these state-level efforts. This model policy affirms that state CSAM laws apply to deepfake CSAM and are punishable under state law irrespective of federal law, providing definitions for terms like "deepfake," "distribute," and "minor".[14]

The proliferation of varied state laws, while demonstrating proactive engagement with the AIG-CSAM threat, creates a complex compliance and enforcement matrix. Investigators and prosecutors, particularly those in federal task forces like CALC units, face a patchwork of definitions, prohibited acts, and standards of proof. This fragmentation can hinder multi-state investigations, create jurisdictional challenges, and lead to inconsistent application of the law in protecting children. Such complexity may eventually fuel calls for federal legislative harmonization or preemption to establish a consistent national baseline for addressing AIG-CSAM.

C. Key Legal Cases and Constitutional Considerations

The legal battleground for AIG-CSAM is largely defined by First Amendment jurisprudence concerning free speech and obscenity, particularly the balance between protecting children and safeguarding expression.

- *New York v. Ferber* **(1982):** This landmark Supreme Court case established that child pornography (CSAM) is not protected by the First Amendment. The Court found that the government's compelling interest in safeguarding the physical and psychological well-being of minors justifies laws prohibiting CSAM.[29] This case underpins all CSAM legislation.
- *Ashcroft v. Free Speech Coalition* **(2002):** This pivotal case significantly complicates efforts to criminalize AIG-CSAM that does *not* involve real children. The Supreme Court struck down provisions of the Child Pornography Prevention Act of 1996 that prohibited computer-generated child pornography depicting fictional

children or adults appearing as children. The Court reasoned that if no real child is harmed in the production or depicted, the government's interest in *Ferber* is not implicated, and such "virtual" child pornography is protected speech unless it meets the separate legal standard for obscenity.[29] However, the Court in *Ashcroft* did uphold the prohibition on "computer morphing" of images of real minors into sexually explicit depictions, finding these "implicate the interests of real children".[29] This distinction is critical for AIG-CSAM that uses identifiable children as source material.

- *Stanley v. Georgia* **(1969) and** *Osborne v. Ohio* **(1990):** *Stanley* affirmed a First Amendment right to possess obscene material in the privacy of one's home. *Osborne* clarified that this right does *not* extend to the private possession of CSAM involving real children.[31]
- **The** *Anderegg* **Case (U.S. District Court, 2025; currently on appeal to the 7th Circuit):** This recent case is poised to be highly influential. Mark Anderegg was charged under the federal child obscenity statute (18 U.S.C. § 1466A) for conduct involving AI-generated CSAM that allegedly did *not* depict real children.[31]
  - The district court **dismissed the charge for private possession** of this "virtual" child obscenity, holding that § 1466A was unconstitutional as applied, based on the precedents of *Stanley* and *Free Speech Coalition*.[31]
  - Crucially, the court **did not dismiss the charges against Anderegg for the production and distribution** of the same AI-generated obscene material.[31] The court reasoned that *Stanley*'s protection was focused on possession and had not been extended by the Supreme Court to the production of obscenity.
  - The *Anderegg* case is believed to be the first federal criminal case involving generative AI, CSAM law, and the First Amendment to reach a federal appeals court, making its outcome highly anticipated.[32]

The central First Amendment debate revolves around whether AIG-CSAM that does not depict an identifiable, actual child, and was not trained using actual CSAM, constitutes protected speech.[5] Legal scholar Riana Pfefferkorn suggests that such AIG-CSAM may indeed be protected under current precedent unless it is legally obscene, or if it can be proven to depict an actual, identifiable child, or if its training data included actual abuse imagery.[5]

The "identifiable actual minor" standard, prevalent in federal law and many state statutes [5], creates a potential evidentiary chokepoint for prosecuting certain types of AIG-CSAM. As AI becomes more adept at generating purely novel, synthetic depictions or heavily anonymized alterations of source images, forensically proving a direct link to an "identifiable actual minor" becomes increasingly difficult, if not impossible.[2] This could leave a category of harmful AIG-CSAM—clearly abusive in appearance and intent but lacking a provable link to a specific real child victim in the image itself—in a more ambiguous legal position under statutes reliant on this standard. This may lead prosecutors to rely more on obscenity charges (as seen in *Anderegg* [31]) or on newer state laws employing an "appears to be a minor" standard [10], though the latter face more direct constitutional scrutiny under *Ashcroft*.

The distinction drawn in the *Anderegg* case between the (potentially protected) private possession of purely synthetic, obscene AIG-CSAM and the (unprotected) production and distribution of such material [31] could influence offender behavior and, consequently, investigative priorities. If this distinction is upheld, offenders might gravitate towards localized, private generation and consumption of AIG-CSAM, utilizing offline tools [20] to minimize digital trails. This would render detection and intervention more challenging, compelling law enforcement to intensify efforts on tracing distribution networks, identifying the tools and platforms used for production, and disrupting the means of creation (e.g., AI models, fine-tuning datasets) rather than focusing solely on end-user possession of purely synthetic material.

D. Defining "Child" and Elements of Offense (Mens Rea, Actus Reus) in Synthetic CSAM Contexts

The application of traditional criminal law elements to AIG-CSAM requires careful consideration:

- **Defining "Child":** Statutes vary significantly. Some require the depiction of an "actual minor" or an "identifiable child".[5] Others have broadened their scope to include images where the subject merely "appears to be a minor" or consists of "virtually indistinguishable depictions".[10] The core challenge for prosecution under the former is proving the "child" element when the image is entirely synthetic or so heavily altered that no real child can be identified from it.
- *Actus Reus* **(Criminal Act):** This refers to the prohibited conduct, such as production, distribution, possession, creation, or alteration, as defined by the specific statute.[1] In the context of AIG-CSAM, the *actus reus* of "production" now encompasses actions like using AI generation tools, inputting prompts into AI models, or fine-tuning algorithms to create illicit content. The *Anderegg* case affirmed that the act of production itself, even of synthetic obscene material, can be prosecuted.[31]
- *Mens Rea* **(Criminal Intent):** Most CSAM statutes require a "knowing" state of mind regarding the prohibited conduct (e.g., knowingly possessing or distributing CSAM).[14] For AIG-CSAM, a key question is the object of this knowledge. Does the offender need to know that the material depicts a *real* child, or is it sufficient that they know it depicts what *appears to be* a child engaged in proscribed sexual conduct and that the material itself is illegal? State laws like Alabama's, which targets "virtually indistinguishable depictions" [28], and California's, which includes matter "digitally altered or generated by AI" [28], suggest a focus on the offender's knowledge of the nature of the image (its visual characteristics and potentially its synthetic origin) and its illicit content, rather than necessarily requiring knowledge of a specific, real child victim within the synthetic image. The *Anderegg* proceedings imply that for the production of obscene AIG-CSAM, the intent to create images that constitute child obscenity is the critical *mens rea* element, even if no real child is ultimately depicted.[31]

Table 2: Comparative Overview of U.S. Federal and Select State Legislative Approaches to AI-Generated CSAM

Jurisdiction	Relevant Statute(s) / Bill	Definition/ Coverage of AIG-CSAM	Key Prohibitions Addressed	"Identifiable Child" vs. "Appears to be Minor" Focus	Penalties	Effective Date / Status
Federal	18 U.S.C. § 2251 et seq.	Includes "computer-generated images indistinguishable from an actual minor, and images created...ap	Production, distribution, possession, receipt, etc. [1]	Primarily focuses on "identifiable, actual minor."	Existing CSAM penalties apply.	Existing law; FBI affirmation 2024.

		pear to depict an identifiable, actual minor." FBI affirms AIG-CSAM is CSAM. [5]				
Federal	TAKE IT DOWN Act (S. 146 / H.R. 633)	AI-generated nonconsensual intimate imagery (NCII). [26]	Nonconsensual disclosure; mandates platform removal.	Focus on nonconsensual intimate imagery, not exclusively CSAM, but relevant to deepfake aspects.	Not specified (likely civil remedies/fines).	Passed Senate Feb 2025; Referred to House Committee.
California	AB 1831, SB 1381 (amending Penal Code)	Expands child pornography statutes to include matter "digitally altered or generated by the use of AI." [28]	Creation, sale, possession, distribution.	Depicts a "person under 18 engaging in or simulating sexual conduct." (Broader than only identifiable actual child)	Existing criminal penalties apply.	January 1, 2025.
Alabama	Alabama Child Protection Act of 2024 (HB 168)	Includes "virtually indistinguishable depictions" created by digital, computer-generated, or other means. [28]	Expands scope of CSAM definition.	Focus on "virtually indistinguishable" from an actual minor.	Existing criminal penalties apply.	October 1, 2024.

ALEC Model	Stop Deepfake CSAM Act (Model Policy)	"Deepfake" defined as visual media manipulated to falsely appear authentic, depicting a minor in sexual conduct. [14]	Knowing possession or distribution; commercial distribution.	Depicting a "minor" (under 18) engaging in sexual conduct; "Recognizable physical characteristics" (actual minor's face/likeness).	State's sentencing guidelines.	Model policy finalized Dec 2023 for state adoption.
38 States	Various state laws (as of Apr 2025)	Vary: some specific to AI, others broader "computer-generated/edited." [10]	Typically possession, production, distribution.	Varies: some require identifiable child, others "appears to be a minor." [10]	Varies by state.	Many enacted in 2024 or 2025.

This table provides a snapshot and is not exhaustive of all state laws or federal proposals. Professionals should consult specific statutory language for definitive guidance.

This comparative overview underscores the dynamic and somewhat fragmented legal landscape. For CALC units and researchers, understanding these variations is crucial for navigating multi-jurisdictional investigations, assessing the applicability of different statutes to specific AIG-CSAM typologies, and analyzing the efficacy and potential constitutional vulnerabilities of diverse legislative strategies.

V. Technological Dimensions: The Creation, Detection, and Attribution of Deepfake CSAM

The technological underpinnings of AIG-CSAM are central to understanding its proliferation and the challenges in combating it. This section details the methods of creation, the technologies for detection and forensic analysis, and the complex issues surrounding attribution.

A. Methods of Deepfake CSAM Creation

The creation of AIG-CSAM relies on increasingly sophisticated and accessible AI technologies:

- **Generative AI Models:** The primary engines are advanced machine learning algorithms, historically Generative Adversarial Networks (GANs), and more recently, diffusion models such as Stable Diffusion.[6] These models learn from vast datasets to generate novel content.
- **Open-Source AI Models:** A critical factor in the proliferation of AIG-CSAM is the availability of powerful open-source models. Tools like **Checkpoint** (base models) and **LoRA (Low-Rank Adaptation)** models (for fine-tuning) can be freely downloaded, modified, and trained on specific datasets—including, illicitly, on CSAM or images of particular children—to produce highly customized AIG-CSAM.[6] This accessibility bypasses

safeguards that might be present in proprietary, closed systems. LoRA models, in particular, can be fine-tuned to depict known victims of child sexual abuse.[7]

- **"Nudify" Applications:** These are specialized AI applications designed to digitally remove clothing from images. They are often used to transform innocent photographs of children into sexually explicit AIG-CSAM.[9]
- **Generation Techniques:** Common methods include face swapping (transplanting one person's face onto another's body in an image or video), puppet-master techniques (animating a static image using the movements of a source video), and complete synthetic generation of individuals and scenes from text prompts or image inputs.[15]
- **Offline Generation Capability:** A significant operational challenge is that many open-source AI models can be downloaded and run on local devices entirely offline.[7] This allows perpetrators to create AIG-CSAM without leaving an immediate online trace, making detection prior to distribution extremely difficult.

The proliferation of open-source models creates a "hydra effect" for enforcement. When models can be easily downloaded, modified, and used privately offline, shutting down a single platform or toolset for AIG-CSAM generation is unlikely to be a lasting solution. New tools can quickly emerge, or offenders can simply shift to entirely private, undetectable generation methods.[6] This resilience means that a purely platform-based takedown strategy is insufficient. Enforcement efforts must also target the dissemination of skills, the datasets used for fine-tuning, and the underlying infrastructure supporting these activities, rather than focusing solely on specific, publicly available tools.

B. Detection Technologies and Forensic Analysis

The detection of AIG-CSAM and the forensic analysis of synthetic media are complex and rapidly evolving fields, characterized by an ongoing "arms race" between generation and detection capabilities.[8]

- **Challenges in Detection:**
  - **Increasing Realism:** The primary challenge is the ever-improving sophistication and realism of deepfakes, making them difficult to distinguish from authentic media, even for trained human eyes.[3]
  - **Evolving Generation Techniques:** As AI generation methods advance, detection algorithms require constant retraining and updates to remain effective.[15]
  - **Real-World Performance:** Detection algorithms that perform well on controlled, academic datasets may exhibit significantly poorer performance when applied to diverse, "in-the-wild" content encountered in actual investigations.[19]
  - **Adversarial Attacks:** Perpetrators can employ adversarial techniques—subtle manipulations of the AI-generated content specifically designed to deceive detection systems.[15]
  - **"Reverse Fakes":** A new forensic challenge involves real CSAM images being intentionally manipulated using AI (e.g., with Stable Diffusion models) to introduce synthetic characteristics, effectively disguising them as AI-generated content to hinder investigations or evade detection as original abuse material.[33]
- **Forensic Tools and Techniques:**
  - **Commercial Forensic Software:**
    - **Amped Authenticate:** This tool is designed for image and video authentication. It examines file structure, metadata, compression schemas, and image content to help differentiate between original camera captures and AI-generated images. It can also utilize Photo Response Non-Uniformity (PRNU) analysis, a technique that can potentially link an image or video to the specific camera sensor that captured it, if a real camera was involved in creating source material.[33]

        - **Magnet Verify:** This tool focuses on video authentication, helping investigators assess the trustworthiness of digital video evidence. It aims to establish whether a video has been edited or modified and to distinguish original camera footage from synthetically produced media. It is designed to generate detailed, legally compliant reports and integrates with other forensic platforms like Magnet Axiom.[19]
    - **General Digital Forensic Methods:** These include meticulous examination of file metadata (though often stripped or altered), analysis of file structures, identification of pixel-level inconsistencies or artifacts common in some AI generation processes, and analysis of image or video compression artifacts.[8]
    - **Machine Learning-Based Detectors:** Many detection approaches utilize machine learning, particularly Convolutional Neural Networks (CNNs) such as ResNet-50, Inception V3, and VGG-16. These models are trained on datasets of real and fake media to learn to identify subtle artifacts or statistical patterns indicative of AI generation.[15]
    - **Multi-Modal Large Language Models (LLMs):** Emerging research is exploring the potential of multi-modal LLMs (e.g., GPT-4V, Gemini variants, Llama Vision) for deepfake detection. The hypothesis is that these models, capable of processing and reasoning about both visual and textual information, might offer advantages by incorporating contextual understanding beyond pixel-level analysis.[35] However, this research is still in its early stages.
- **Explainable AI (XAI) in Detection:**
    - A significant limitation of many advanced AI detection models is their "black box" nature—they can classify media as real or fake with some accuracy, but the internal reasoning for their decisions is often opaque.[15] This lack of transparency is a major hurdle for building trust in these tools and for their admissibility as evidence in legal proceedings, where the basis for an expert opinion must often be clearly articulated.
    - XAI techniques aim to make the decision-making processes of AI models more interpretable. For deepfake detection, this is crucial.[15]
        - The **Network Dissection Algorithm**, when applied to CNNs like ResNet-50, Inception V3, and VGG-16, can help identify which specific visual features (e.g., particular facial regions or artifacts) the model is focusing on when classifying an image as real or fake.[36]
        - **LIME (Local Interpretable Model-Agnostic Explanations)** is another technique used to provide localized explanations for the predictions of complex models, enhancing their validity and reliability.[37]
    - By providing insights into *how* a detection model arrives at its conclusion, XAI can help in debugging and improving models, identifying potential biases in training data, increasing robustness against adversarial attacks, and bolstering the confidence of law enforcement personnel and courts in utilizing these technologies.[15] The "explainability gap" is thus a critical area of research and development.
-

The problem of authenticating digital media is twofold. It is not merely about identifying AIG-CSAM as synthetic; it is equally about reliably authenticating genuine media to prove that it *has not* been tampered with or AI-generated, especially when such claims are made by defendants in court.[19] The very existence and proliferation of sophisticated deepfake technology can lead to a "liar's dividend," where the authenticity of *all* digital evidence is cast into doubt, potentially undermining the fact-finding process in legal proceedings.[34]

C. Attribution Challenges and Emerging Solutions

Attributing AIG-CSAM to its specific creators or distributors is one of the most formidable challenges:

- **Anonymity Tools:** Perpetrators frequently leverage encryption, anonymous networks like Tor (often associated with the Dark Web), virtual private networks (VPNs), and offshore hosting platforms to obscure their identities and locations.[1]

- **Complex Digital Trails:** The process of creating and disseminating AIG-CSAM can involve multiple layers of editing, file transformations, and re-uploads across various platforms. Social media platforms, for instance, often re-encode uploaded media, which can strip original metadata and alter file characteristics, complicating efforts to trace content back to its original source with forensic certainty.[19]
- **Technical Solutions for Provenance and Attribution (Emerging and In-Development):**
  - **Digital Watermarking:** This involves embedding a perceptible (visible) or imperceptible (invisible) signal or pattern into AI-generated content. The watermark could theoretically carry information about the AI model used, the date of creation, or simply indicate that the content is AI-generated.[34]
    - *Limitations:* Watermarks, especially invisible ones, can be vulnerable to removal or degradation through image/video processing (e.g., compression, cropping, filtering). There is no universal standard for watermarking AI content, and open-source models or custom-built tools used by offenders are unlikely to incorporate them. Furthermore, inconsistent international laws regarding mandatory watermarking mean that content can circulate globally without such labels.[38]
  - **Digital Fingerprinting:** This technique generates a unique identifier (a "fingerprint" or hash) based on the intrinsic characteristics of the content itself. This fingerprint can then be registered in a database with associated metadata (e.g., confirming it as AI-generated by a specific model). When new content is encountered, its fingerprint can be compared against the database.[38]
    - *Limitations:* Fingerprints can be sensitive to even minor alterations in the content, potentially leading to mismatches. The effectiveness relies on comprehensive databases and widespread adoption of fingerprinting by AI content generators.
  - **Cryptographic Metadata and Content Provenance Initiatives:** Efforts like the Coalition for Content Provenance and Authenticity (C2PA) are working to establish technical standards for securely binding metadata to digital content, creating a verifiable record of its origin and modification history.[34] This "provenance" data would be cryptographically signed to ensure its integrity.
    - *Limitations:* Metadata can still be stripped from files, intentionally or unintentionally. Widespread adoption by all content creation tools and platforms is necessary for broad effectiveness.
  - **Blockchain-Based Content Authentication:** The use of blockchain technology has been proposed as a means to create immutable, transparent records of content creation and modification, thereby enhancing authentication and traceability.[8]
  - **Analysis of AI Model Artifacts:** Research is underway to determine if specific AI generation models or techniques leave subtle, unique artifacts or statistical signatures in the output they produce. Identifying such signatures (e.g., linked to specific Checkpoint or LoRA models as suggested by analysis of dark web forum discussions [7]) could potentially help attribute AIG-CSAM to a particular class of generation tools, even if not to a specific user without further corroborating evidence.

Even if robust technical solutions for watermarking or provenance are developed, their ultimate effectiveness hinges on widespread, standardized adoption by AI model creators, software developers, and content platforms globally. This is not merely a technical challenge but also a significant hurdle related to international cooperation, industry self-regulation, and potentially, legislative mandates. Without global consensus and consistent implementation, particularly by developers of open-source tools that are easily accessible to malicious actors, these solutions will have limited impact on the AIG-CSAM problem.[38]

Table 3: Overview of Deepfake CSAM Detection, Authentication, and Attribution Technologies

Technology/Method	**Mechanism**	**Primary Use(s)**	**Strengths**	**Limitations/Challenges**
Forensic Software (e.g., Amped Authenticate, Magnet Verify)	Analysis of file structure, metadata, compression, pixel data, PRNU (photo response non-uniformity), video stream integrity. [19]	Detection, Authentication	Comprehensive analysis, PRNU for camera linking, legally compliant reporting (some tools).	Requires expert use, evolving with AI; PRNU depends on original camera data.
Machine Learning (ML) Detectors (e.g., CNNs)	Trained on real/fake datasets to identify AI-generated artifacts/patterns. [15]	Detection	Can automate detection of some fakes at scale.	"Black box" nature, vulnerable to adversarial attacks, performance drop on unseen fakes, requires constant retraining. [15]
Explainable AI (XAI) (e.g., Network Dissection, LIME)	Provides insights into ML model decision-making (e.g., which features are used for classification). [15]	Detection (Interpretability)	Increases trust, aids debugging, may support legal admissibility by explaining findings. [15]	Still an emerging field, complexity, ensuring explanations are truly reflective of model reasoning.
Digital Watermarking	Embedding visible/invisible signals in content indicating AI origin or source. [34]	Attribution (Potential), Authentication	Simple concept, can be integrated by compliant AI generators.	Easily stripped/bypassed (especially by malicious actors), no universal standard, not used by open-source tools, international inconsistency. [38]
Digital Fingerprinting	Generating unique content-based codes linked to origin databases.	Attribution (Potential), Authentication	Can identify known AI-generated content if	Sensitive to alterations, relies on comprehensive databases and

	[38]		fingerprinted.	adoption. [38]
Cryptographic Metadata / Content Provenance (e.g., C2PA)	Securely embedding verifiable data about content origin and modification history. [34]	Authentication, Attribution (Potential)	Provides verifiable chain of custody if adopted widely.	Metadata can be stripped, relies on universal adoption by creation tools/platforms. [38]
Blockchain-based Authentication	Using distributed ledger technology for immutable records of content creation/modification. [8]	Authentication, Attribution (Potential)	Tamper-resistant, transparent record-keeping.	Scalability, cost, complexity of implementation, widespread adoption needed.
Analysis of AI Model Artifacts	Identifying unique signatures left by specific AI generation models/techniques. [7]	Attribution (Tool-level)	May link content to a class of AI tools if specific artifacts are robustly identifiable.	Research intensive, requires deep understanding of AI models, may not identify individual users.

This table offers a structured overview of the current technological landscape. For professionals, it underscores that while various tools and methods exist, each has inherent limitations. No single technology provides a complete solution, emphasizing the need for a multi-layered investigative and analytical approach, continuous research and development, and a critical understanding of the evidentiary value and constraints of each technique.

VI. Operational Challenges for Law Enforcement and CALC Units

The proliferation of AIG-CSAM presents formidable operational challenges for law enforcement agencies (LEAs) and specialized units such as Child Abuse, Internet Crimes Against Children (ICAC) Task Forces, and Cybercrime (CALC) units. These challenges span the investigative, forensic, and resource domains.

A. Investigative Hurdles

- **Overwhelming Volume of Material:** The sheer quantity of CSAM, already a massive problem, is now significantly amplified by the ease with which AIG-CSAM can be generated and disseminated.[3] NCMEC's receipt of over 100 million suspected CSAM reports in 2023, even before the full impact of AIG-CSAM was felt, indicates the scale of the data deluge.[3] This volume far exceeds existing investigative capacities.
- **Sophistication and Realism of Fakes:** Distinguishing AIG-CSAM from authentic CSAM, or identifying subtle AI manipulations within genuine CSAM, requires specialized tools, advanced training, and considerable time. The increasing hyper-realism of synthetic media makes this differentiation progressively more difficult, even for experienced analysts.[3]
- **Anonymity and Encryption:** Perpetrators routinely employ technologies and tactics to anonymize their

activities, including the use of anonymous networks like the Dark Web, robust encryption methods, and globally distributed infrastructure. These measures make tracing the origin of AIG-CSAM and attributing it to specific individuals exceedingly challenging.[1]

- **Offline Generation:** The capability to download and run AI models offline to generate CSAM means that such material can be created without leaving an immediate online footprint for platforms or LEAs to detect prior to its distribution.[7] This clandestine creation process significantly hinders proactive detection.
- **The Attribution Puzzle:** Even when AIG-CSAM is identified, proving who created the specific content, meticulously tracking its distribution pathways, and establishing legal jurisdiction, especially when content crosses international borders, remain major investigative obstacles.[8]

The confluence of these factors creates a critically complex investigative triage situation for law enforcement. With AIG-CSAM flooding reporting systems, investigators face the daunting task of prioritizing cases. They must decide whether to focus resources on images that *might* depict a real child currently suffering abuse, or on those that are clearly synthetic yet still illegal and harmful. This decision-making process is severely complicated by the increasing difficulty in distinguishing between authentic and synthetic media.[3] Any mis-prioritization in this high-stakes environment could tragically delay the identification and rescue of real child victims, as time spent analyzing purely synthetic material is time diverted from potentially life-saving interventions.[19]

B. Forensic and Evidentiary Challenges in Prosecution

Successfully prosecuting cases involving AIG-CSAM hinges on overcoming significant forensic and evidentiary hurdles:

- **Authenticating Digital Evidence:** A fundamental requirement in legal proceedings is the authentication of evidence—proving that a piece of digital media is what it purports to be. The advent of sophisticated AI fakes complicates this process immensely. Prosecutors must be able to establish the authenticity and reliability of digital media, whether it is proving that an image *is* AIG-CSAM or, conversely, proving that an image alleged by the defense to be a fake is, in fact, genuine CSAM.[19]
- **Admissibility of Detection Tool Findings:** The results generated by deepfake detection tools may face challenges to their legal admissibility if the underlying methodology is not transparent, scientifically validated, and robustly vetted.[19] The "black box" nature of some AI-driven detection tools, where the reasoning behind a classification is not easily explainable, can be a significant barrier to their acceptance in court.[15]
- **Proving "Identifiable Actual Minor":** If the applicable statute requires proving that AIG-CSAM depicts an "identifiable actual minor," this can be forensically impossible for highly altered images or those that are purely synthetic and do not incorporate the likeness of any real child.[5]
- **Chain of Custody:** Maintaining an unbroken and verifiable chain of custody for digital evidence is crucial for its admissibility. This is already complex for digital files and becomes more so with AIG-CSAM that may have been stored in the cloud, shared across multiple platforms (each potentially altering the file), or generated using ephemeral tools.[19]

The ease with which AIG-CSAM can be created also contributes to a "digital dust" problem for investigators. Law enforcement may be inundated with vast quantities of AIG-CSAM that, while illegal, is forensically "sterile." This means the material itself offers few, if any, leads back to an identifiable real-world victim needing rescue or to a traceable producer, particularly if it was generated offline using open-source tools and distributed anonymously. While the possession and distribution of such material remain criminal offenses, its investigative value for the primary mission of many CALC units—victim identification and rescue—can be low. Nevertheless, this material still consumes valuable analytical resources, diverting them from leads that might be more actionable in terms of immediate child protection.[8]

C. Resource Asymmetry and Training Needs

The operational capacity of LEAs is severely strained by the AIG-CSAM phenomenon:

- **Resource Crisis:** There is a profound asymmetry between the ease and speed with which offenders can produce vast quantities of AIG-CSAM and the time-consuming, resource-intensive nature of investigating these offenses.[8] A single individual can generate hundreds or thousands of synthetic images, while each investigation demands significant technical expertise and man-hours.
- **Specialized Training Requirements:** Investigators, forensic analysts, and prosecutors require continuous, specialized training to keep pace with rapidly evolving AI generation techniques, new detection tools, digital forensic methodologies for synthetic media, and the shifting legal landscape surrounding AIG-CSAM.[8]
- **Lack of Advanced Tools and Expertise:** Many LEAs, particularly smaller or less-resourced agencies, currently lack access to the necessary AI-driven detection tools, advanced forensic software, and in-house expertise required to effectively investigate AIG-CSAM.[2]
- **International Cooperation:** Given the inherently global and borderless nature of online crime and CSAM distribution, effective international cooperation among LEAs is essential. However, such cooperation can be hampered by differing legal systems, procedural delays, and resource disparities across jurisdictions.[8]

The constant exposure to CSAM, whether authentic or synthetic, inflicts vicarious trauma on investigators. The added frustrations of dealing with technologically advanced, difficult-to-trace AIG-CSAM, the sheer volume of such material, and the awareness of its increasing ease of generation, could significantly exacerbate stress, burnout, and impact the psychological well-being and retention of these highly specialized professionals. The feeling of being locked in a perpetual "arms race" [15] against an ever-evolving technological threat, where successes may feel incremental against a tide of new material, can be profoundly demoralizing.

VII. The Human Element: Psychological and Societal Impact of Deepfake CSAM

The proliferation of AIG-CSAM inflicts profound psychological harm on individuals and carries broader, detrimental societal implications. Understanding these human costs is essential for formulating effective responses.

A. Victimology: Trauma, Re-victimization, and the "Reality" of Synthetic Abuse

The impact on children whose likenesses are used in AIG-CSAM, or who are targeted by it, is severe, irrespective of the "synthetic" nature of the material itself.

- **Psychological Impact on Children Depicted:**
  - Children whose innocent images are transformed into AIG-CSAM, or who are depicted synthetically, experience intense **humiliation, shame, anger, feelings of violation, and self-blame**.[11] These reactions are not diminished by the knowledge that the depicted acts did not physically occur as shown.
  - The experience can lead to **immediate and ongoing emotional distress, social withdrawal from family and school, and profound challenges in forming and sustaining trusting relationships**.[11]
  - If deepfake CSAM circulates within a child's school community or peer groups, the victim is often subjected to **bullying, teasing, and harassment**. Each instance of sharing or viewing amplifies the trauma.[11]
  - Significant **harm to reputation, diminished academic performance, and a loss of confidence about the future** can result, driven by the fear that these fake but explicit images will remain permanently

accessible online.[11]

  - A particularly insidious aspect is the **fear of not being believed** by parents, peers, or authorities. Victims may worry that others will believe the deepfake is real, or conversely, that their genuine distress will be dismissed because the image is "only a fake".[3] This "believability burden" creates a complex psychological trap, intensifying barriers to seeking help and support. Research indicates that boys, in particular, are less likely to disclose victimization by deepfake pornography or CSAM.[11]
- **Re-victimization of Existing Survivors:**
  - AIG-CSAM is actively used to create new abusive material featuring children who are already survivors of sexual abuse. Malicious actors fine-tune AI models using existing CSAM or publicly available images of known victims, thereby generating novel depictions of their abuse. This constitutes a profound re-victimization, compounding their original trauma and perpetuating their suffering.[3]
- **Sextortion Using AIG-CSAM:**
  - The use of AIG-CSAM in sextortion schemes is a rapidly growing concern.[3] Perpetrators create explicit deepfakes of a child, often using innocent photos sourced from social media, and then blackmail the child or their family for money or further material.
  - For the victims of such schemes, typically adolescent boys aged 14-17 [11], the fact that the explicit image is AI-generated does not lessen the terror, violation, or fear of exposure. The psychological leverage is immense because the content is both deeply personal and humiliating, while the perpetrator might attempt to deny full culpability by claiming it's "just AI."
- **The "Fake" Label's Irrelevance to Harm:**
  - Experts and victim advocates consistently emphasize that the harm inflicted by AIG-CSAM is very real, regardless of the synthetic origin of the image itself.[3] The emotional, psychological, and reputational damage is tangible.
  - However, there is a concerning perception gap among some young people. Research by Thorn indicated that while many youths recognize the harm, around 1 in 6 believed deepfake nudes were either not harmful or that harm depended on the situation, with top reasons being that the imagery was "fake" or "not real," or involved no physical harm.[21] This highlights a dangerous misunderstanding of the nature of this abuse and the need for targeted education.

B. Broader Societal Implications

The rise of AIG-CSAM extends beyond individual harm, posing wider societal challenges:

- **Normalization and Desensitization:** The increasing prevalence of AIG-CSAM, alongside other forms of deepfakes and synthetic media, risks normalizing the sexualization of minors online and desensitizing the public to the severe distress such material causes.[2] This "normalization of SOM (sexualization of minors) online" [11] can erode protective societal norms.
- **Erosion of Trust in Digital Media:** The ease with which convincing fake images and videos can be created undermines general trust in all forms of visual media.[8] This "liar's dividend" can complicate the use of genuine digital evidence in legal proceedings and pollute public discourse, making it harder to discern truth from fabrication.
- **Increased Demand for CSAM:** The ability of AI to generate "novel" and highly customized CSAM could potentially fuel existing demand among offenders and even create new markets for specific types of abusive content, thereby perpetuating the cycle of exploitation.[1]
- **Impact on Child Development:** Exposure to, or victimization by, AIG-CSAM can have deleterious long-term effects on children's psychological development, their body image, their understanding of healthy sexuality

and relationships, and their sense of safety in online environments.[11] If young people are increasingly exposed to AI-generated explicit content, it may warp their understanding of real-world sexual interactions, the critical importance of consent (which is entirely absent in CSAM), and the value of authenticity in human connections.

VIII. Strategic Imperatives: Policy Recommendations and Collaborative Pathways Forward

Confronting the complex and rapidly evolving threat of AIG-CSAM requires a multi-pronged strategic approach, encompassing robust legal and regulatory reforms, enhanced technological countermeasures, strengthened multi-stakeholder cooperation, and comprehensive public awareness initiatives.

A. Strengthening Legal and Regulatory Safeguards

A clear and consistently applied legal framework is paramount.

- **NCMEC Recommendations:**
  - Federal and state laws must be updated to explicitly clarify that AIG-CSAM is illegal, regardless of whether a "real" child is depicted in the final image, if it appears to be a child.[18]
  - Civil remedies should be available for victims of AIG-CSAM, allowing them to seek recourse for the harm suffered.[18]
  - Platforms developing or deploying generative AI must bear greater responsibility. This includes incorporating "safety by design" principles from the outset, ensuring AI models are not trained on CSAM datasets, actively teaching AI models *not* to create CSAM, implementing effective systems to detect, report, and remove attempts to generate CSAM, and holding platforms accountable for misuse of their tools.[18] This call for "safety by design" represents a fundamental paradigm shift, moving responsibility "upstream" to the creators of AI technology, rather than relying solely on reactive content moderation. It implies embedding ethical considerations and preventative measures directly into the AI research, development, and deployment lifecycle, which may involve trade-offs with innovation speed or model capabilities.
- **UNICRI Recommendations for Governments:**
  - Legislation should be revised to explicitly criminalize the creation, possession, and distribution of all forms of AIG-CSAM.[23]
  - Legal accountability should extend to service providers hosting AIG-CSAM and AI developers whose models facilitate its creation.[23]
  - Systemic reforms in internet and social media governance are needed, moving beyond reactionary legislation to address underlying structural issues.[23]
  - Governments should foster partnerships with technology companies to ensure accountability and the implementation of robust child safeguards, while also investing heavily in law enforcement technology, education, and training.[23]
  - A notable recommendation is to treat the high rates of CSAM offending as a **public health issue**, investing in solutions for perpetrators who seek help, suggesting a more holistic, preventative, and rehabilitative approach alongside traditional enforcement.[23] This long-term strategy complements immediate law enforcement actions by aiming to reduce the pool of offenders over time through understanding root causes, risk factors, and implementing preventative interventions.
- **SaferAI for Children Coalition Recommendations:**
  - Advocacy for comprehensive legislative and regulatory frameworks that establish an effective,

    whole-of-system approach to child safety in the age of AI.[13]
  - Continuous monitoring of existing laws to identify gaps and promote international cooperation to harmonize legal frameworks and enforcement mechanisms across borders.[13]
- **Addressing Constitutional and Free Speech Concerns:**
  - Any new legislation must be carefully crafted to be narrowly targeted and constitutionally sound, particularly in light of First Amendment protections for speech and the precedents set in cases like *Ashcroft v. Free Speech Coalition*.[30] The ongoing debate surrounding the TAKE IT DOWN Act highlights the enduring tension between robust victim protection measures and the safeguarding of free expression and user privacy, especially concerning platform liability and the risks of overbroad automated content moderation.[27]

B. Enhancing Technological Countermeasures and Research

Technological innovation is crucial for both detecting AIG-CSAM and preventing its creation and spread.

- **Investment in Detection and Authentication R&D:** Sustained investment is needed for the continuous development of robust, explainable, and legally admissible tools for identifying AIG-CSAM and, equally importantly, for authenticating genuine media.[8]
- **Support for Explainable AI (XAI) Development:** Research into XAI techniques must be prioritized to make AI detection tools transparent, interpretable, and trustworthy for use in legal and operational settings.[15] This is critical for overcoming the "black box" problem and ensuring that findings from AI tools can be effectively presented and scrutinized in court.
- **Watermarking and Digital Provenance Standards:** Efforts to develop and encourage the widespread adoption of effective, robust, and standardized watermarking and content provenance technologies should be supported.[23] However, the limitations of these technologies, particularly their vulnerability to removal and the need for universal adoption, must be acknowledged.[38]
- **AI-Driven Solutions for Proactive Protection:** As advocated by the SaferAI for Children Coalition, AI itself can be leveraged for child protection. This includes developing AI-powered tools to detect and prevent Online Child Sexual Exploitation (OCSE), automate the initial identification and categorization of CSAM for law enforcement (reducing human exposure), and deploy AI to detect and report known, new, or livestreamed CSAM—potentially even within encrypted environments—and to detect and block CSAM creation, storage, and distribution at the device level.[13]

C. Fostering Multi-Stakeholder Cooperation

No single entity can adequately address the AIG-CSAM crisis. Effective responses require deep and sustained collaboration.

- **Broad Coalition Needed:** NCMEC, UNICRI, and the SaferAI for Children Coalition all emphasize the necessity of GAI technology creators, legislators, child-serving professionals, law enforcement agencies, the broader private sector, academia, and NGOs working in concert.[13]
- **Private Sector Responsibility:**
  - **AI Developers:** Must implement "safety by design" from the earliest stages of model development, rigorously scrutinize training datasets to exclude CSAM, and invest in technologies like watermarking or content filtering capabilities within their models.[18]
  - **Technology Platforms (Social Media, Hosting Providers, etc.):** Must update algorithms to prevent the

recommendation of explicit content and limit interactions that could facilitate grooming. They need to engage in robust content moderation to actively remove AIG-CSAM, block websites known to host CSAM, and potentially refuse to advertise on platforms that fail to adequately monitor and remove such material.[23]

- **International Cooperation for Law Enforcement:** Given the transnational nature of AIG-CSAM creation and distribution, enhanced international cooperation is vital for sharing intelligence on emerging software, perpetrator tactics, and investigative best practices.[8]

Effective multi-stakeholder collaboration, however, requires overcoming inherent challenges such as misaligned incentives and historical trust deficits between different sectors. For instance, technology companies may prioritize rapid innovation and minimal content liability, while law enforcement requires access to data and platform accountability. Building sustained, impactful collaboration necessitates mechanisms that align these varied interests—such as clear safe harbor provisions for good faith cooperation, public-private funding for safety-focused R&D, and the establishment of transparent processes and shared, measurable goals that transcend public relations statements.

D. Public Awareness and Education Initiatives

Education is a critical pillar in preventing the creation and mitigating the impact of AIG-CSAM.

- **Youth Education:** Young people need to be educated about the safe and ethical use of generative AI, the severe legal and personal consequences of misusing these tools to create or share explicit images of minors (including peers), and the nature of online harms like sextortion.[18]
- **Caregiver Education:** Parents, guardians, and educators must be equipped with knowledge about the evolving online dangers facing children, including the specific threats posed by AIG-CSAM.[23] UNICRI recommends that caregivers:
  - Reconsider the extent to which they post images of children on social media, given the risk of misuse.
  - Stay informed about the latest technologies and how they can be exploited.
  - Engage in open and frequent conversations with children about online dangers, including the potential for their images to be altered and the permanence of online content.
  - Clearly communicate to teenagers that "nudify" apps and similar tools that generate AIG-CSAM are illegal and harmful.
  - Educate children about the risks of sextortion and ensure they know how to respond and seek help.
  - Utilize existing educational toolkits and advice from reputable child online safety organizations and disseminate this information within their communities.
- **Broader Digital Literacy Programs:** Public awareness campaigns are needed to enhance general understanding of deepfake technology, the risks associated with AI-generated content, and fundamental online safety practices.[2]

IX. Conclusion: Confronting the Deepfake CSAM Crisis and Protecting Children in the Digital Age

A. Recapitulation of Key Findings

The emergence of AI-generated Child Sexual Abuse Material represents a significant and rapidly escalating escalation in the landscape of online child exploitation. This report has detailed how AIG-CSAM, facilitated by increasingly accessible and sophisticated artificial intelligence technologies, poses multifaceted threats. These include the direct psychological trauma to children whose likenesses are used (whether real or synthetically derived from innocent images), the re-victimization of existing survivors, the overwhelming of law enforcement and child protection systems with vast new quantities of illicit material, and the creation of complex legal and evidentiary challenges. The very nature of AIG-CSAM—often hyper-realistic and difficult to distinguish from authentic abuse imagery—strains investigative resources and complicates prosecutorial efforts. The legal and technological landscapes are currently in a reactive posture, struggling to keep pace with the speed of AI advancements and their malicious applications.

B. The Urgency of Coordinated Action

The analysis presented underscores a critical reality: no single entity, legislative act, or technological solution can independently resolve the AIG-CSAM crisis. Confronting this challenge effectively demands a unified, proactive, and adaptive strategy that integrates the efforts of governments, the technology industry (from AI developers to content platforms), law enforcement agencies at all levels, academic researchers, non-governmental child protection organizations, educators, and caregivers. The interconnectedness of the technological, legal, operational, and human dimensions of this problem necessitates a holistic and collaborative response.

The AIG-CSAM crisis serves as a stark bellwether for broader AI governance challenges. The manner in which society, legal systems, and technological safeguards adapt to combat AIG-CSAM will invariably inform and potentially set precedents for addressing other malicious uses of artificial intelligence, such as sophisticated disinformation campaigns, AI-enabled financial fraud, and the ethical dilemmas posed by autonomous systems. Successfully tackling AIG-CSAM requires the development and implementation of frameworks for AI ethics, accountability, and safety by design—principles that will have far-reaching implications beyond this specific form of abuse.

C. Future Outlook and Emerging Concerns

The trajectory of AI development suggests that the capabilities for generating synthetic media will continue to advance in realism, ease of use, and accessibility.[2] This implies an ongoing "cat and mouse" dynamic between those who create malicious deepfakes and those who work to detect and mitigate them.[15] Future concerns may include the automated mass-production of AIG-CSAM, its integration with other forms of cybercrime (e.g., more convincing phishing or grooming operations powered by generative AI), and the potential for AI to be used to create entirely novel forms of child exploitation that are not yet fully conceptualized. The challenge is not static; it is a moving target that will demand continuous vigilance and innovation.

Sustaining long-term vigilance and adaptability is therefore paramount. The fight against AIG-CSAM is not a campaign with a defined endpoint but an ongoing process of societal and technological adaptation. This necessitates a sustained commitment to funding for research and development of countermeasures, robust inter-agency and international cooperation, continuous training for professionals, and an unwavering societal consensus that prioritizes the safety and well-being of children above technological permissiveness when these values come into conflict.

D. Concluding Call to Action for the Professional Audience

This report is intended to equip researchers, CALC unit personnel, and other domain experts with a comprehensive understanding of the AIG-CSAM threat. The path forward requires concerted action:

- **For Academic Researchers:** Continue rigorous investigation into the nuanced aspects of AIG-CSAM, including its psychological impact on diverse victim populations, the efficacy and limitations of detection and attribution technologies, the evolving tactics of offenders, and the socio-legal implications of synthetic media. Longitudinal studies and cross-disciplinary research are particularly needed.

- **For CALC Units and Law Enforcement Professionals:** Advocate for the necessary resources, including specialized training, advanced technological tools, and adequate staffing to manage the increased caseloads and complexities introduced by AIG-CSAM. Share best practices, investigative techniques, and intelligence on emerging threats through established networks and by fostering new collaborative channels. Contribute to the validation and operationalization of new forensic tools.
- **For All Experts in the Field:** Actively contribute to policy development by providing evidence-based insights to legislators and regulatory bodies. Engage in public awareness initiatives to educate communities about the risks of AIG-CSAM and the importance of digital safety. Champion the implementation of robust child safety measures within technological development and deployment, advocating for ethical AI principles that prioritize the protection of vulnerable populations.

The challenge posed by AI-generated Child Sexual Abuse Material is profound, but not insurmountable. Through dedicated research, strategic investment, robust legal frameworks, technological innovation, and unwavering multi-stakeholder collaboration, it is possible to mitigate this threat and enhance the protection of children in an increasingly complex digital age.

Works cited

1. www.justice.gov, accessed May 13, 2025, https://www.justice.gov/d9/2023-06/child_sexual_abuse_material_2.pdf
2. Digital child abuse: Deepfakes and the rising danger of AI-generated exploitation, accessed May 13, 2025, https://lens.monash.edu/@politics-society/2025/02/25/1387341/digital-child-abuse-deepfakes-and-the-rising-danger-of-ai-generated-exploitation
3. Combatting AI-Generated CSAM | Wilson Center, accessed May 13, 2025, https://www.wilsoncenter.org/article/combatting-ai-generated-csam
4. Combatting AI-Generated CSAM - Across Karman - Wilson Center, accessed May 13, 2025, https://acrosskarman.wilsoncenter.org/article/combatting-ai-generated-csam
5. Combatting AI-Generated CSAM - 5G Beyond Borders - Wilson Center, accessed May 13, 2025, https://5g.wilsoncenter.org/article/combatting-ai-generated-csam
6. Combating the rise of AI-generated child sexual abuse material - Humanium, accessed May 13, 2025, https://www.humanium.org/en/combating-the-rise-of-ai-generated-child-sexual-abuse-material/
7. Unveiling AI's Threats to Child Protection: Regulatory efforts to Criminalize AI-Generated CSAM and Emerging Children's Right - arXiv, accessed May 13, 2025, https://www.arxiv.org/pdf/2503.00433
8. The Deepfake Crisis: How AI is Reshaping Criminal Justice | Crime ..., accessed May 13, 2025, https://www.crimetraveller.org/2025/02/the-deepfake-crisis-ai-criminal-justice/
9. Artificial intelligence and child sexual abuse: A rapid evidence assessment - ResearchGate, accessed May 13, 2025, https://www.researchgate.net/profile/Timothy-Cubitt/publication/388275053_Artificial_intelligence_and_child_sexual_abuse_A_rapid_evidence_assessment/links/6791fbfe82501639f5064715/Artificial-intelligence-and-child-sexual-abuse-A-rapid-evidence-assessment.pdf

10. State Laws Criminalizing AI-generated or Computer-Edited CSAM - Enough Abuse, accessed May 13, 2025, https://enoughabuse.org/get-vocal/laws-by-state/state-laws-criminalizing-ai-generated-or-computer-edited-child-sexual-abuse-material-csam/
11. The Impact of Deepfakes, Synthetic Pornography, & Virtual Child Sexual Abuse Material, accessed May 13, 2025, https://www.aap.org/en/patient-care/media-and-children/center-of-excellence-on-social-media-and-youth-mental-health/qa-portal/qa-portal-library/qa-portal-library-questions/the-impact-of-deepfakes-synthetic-pornography--virtual-child-sexual-abuse-material/
12. Are Deepfakes Illegal? Overview Of Deepfake Laws And Regulations - HyperVerge, accessed May 13, 2025, https://hyperverge.co/blog/are-deepfakes-illegal/
13. 26980386.fs1.hubspotusercontent-eu1.net, accessed May 13, 2025, https://26980386.fs1.hubspotusercontent-eu1.net/hubfs/26980386/Website%20Resources/AI%20&%20Child%20Protection%20A%20Collaborative%20Approach%20to%20a%20Safer%20Future%20(SaferAI%20for%20Children%20Coalition)%20(3).pdf
14. Stop Deepfake CSAM Act - American Legislative Exchange Council ..., accessed May 13, 2025, https://alec.org/model-policy/stop-deepfake-csam-act/
15. Generative Artificial Intelligence and the Evolving Challenge of Deepfake Detection: A Systematic Analysis - MDPI, accessed May 13, 2025, https://www.mdpi.com/2224-2708/14/1/17
16. Increasing Threat of DeepFake Identities - Homeland Security, accessed May 13, 2025, https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf
17. CyberTipline Data - MissingKids.org, accessed May 13, 2025, https://www.missingkids.org/cybertiplinedata?preview=true&site_id=816
18. Generative AI CSAM is CSAM - MissingKids.org, accessed May 13, 2025, https://www.missingkids.org/blog/2024/generative-ai-csam-is-csam
19. The challenge of authenticating media in the age of AI-generated CSAM - Magnet Forensics, accessed May 13, 2025, https://www.magnetforensics.com/blog/the-challenge-of-authenticating-media-in-the-age-of-ai-generated-csam/
20. How AI is being abused to create child sexual abuse material ..., accessed May 13, 2025, https://www.iwf.org.uk/about-us/why-we-exist/our-research/how-ai-is-being-abused-to-create-child-sexual-abuse-imagery/
21. Deepfake nudes and young people - Thorn Research, accessed May 13, 2025, https://www.thorn.org/research/library/deepfake-nudes-and-young-people/
22. NCMEC Releases New Data: 2024 in Numbers - MissingKids.org, accessed May 13, 2025, https://www.missingkids.org/blog/2025/ncmec-releases-new-data-2024-in-numbers

23. unicri.it, accessed May 13, 2025, https://unicri.it/sites/default/files/2024-09/Generative-AI-New-Threat-Online-Child-Abuse.pdf
24. New film exposes AI's role in online child sexual exploitation and calls for urgent global action, accessed May 13, 2025, https://www.weprotect.org/news/protect-us-film-launched/
25. Artificial intelligence and child sexual abuse: A rapid evidence assessment - Australian Institute of Criminology, accessed May 13, 2025, https://www.aic.gov.au/sites/default/files/2025-01/ti711_artificial_intelligence_and_child_sexual_abuse.pdf
26. U.S. Tech Legislative & Regulatory Update – First Quarter 2025 ..., accessed May 13, 2025, https://www.insideglobaltech.com/2025/04/23/u-s-tech-legislative-regulatory-update-first-quarter-2025/
27. The TAKE IT DOWN Act: A Flawed Attempt to Protect Victims That ..., accessed May 13, 2025, https://www.eff.org/deeplinks/2025/02/take-it-down-act-flawed-attempt-protect-victims-will-lead-censorship
28. U.S. AI Law Tracker, accessed May 13, 2025, https://ai-law-center.orrick.com/wp-content/uploads/2025/04/Orrick-US-AI-Law-Tracker-2025-03-31.pdf
29. Legal fight against AI-generated child pornography is complicated – a legal scholar explains why, and how the law could catch up - Route Fifty, accessed May 13, 2025, https://www.route-fifty.com/digital-government/2025/02/legal-fight-against-ai-generated-child-pornography-complicated-legal-scholar-explains-why-and-how-law-could-catch/402945/?oref=ge-category-lander-river
30. Addressing Computer-Generated Child Sex Abuse Imagery: Legal ..., accessed May 13, 2025, https://www.lawfaremedia.org/article/addressing-computer-generated-child-sex-abuse-imagery-legal-framework-and-policy-implications
31. Court Rules That Constitution Protects Private Possession of AI ..., accessed May 13, 2025, https://www.techpolicy.press/court-rules-that-constitution-protects-private-possession-of-aigenerated-csam
32. Court Rules That Constitution Protects Private Possession of AI-Generated CSAM, accessed May 13, 2025, https://www.techpolicy.press/court-rules-that-constitution-protects-private-possession-of-aigenerated-csam/
33. Fighting Deepfakes: AI-Generated CSAM and the Tools to Detect It - Amped Blog, accessed May 13, 2025, https://blog.ampedsoftware.com/2025/02/19/fighting-deepfakes-ai-generated-csam-and-the-tools-to-detect-it
34. www.ncsc.org, accessed May 13, 2025, https://www.ncsc.org/__data/assets/pdf_file/0007/111103/Evidentiary-Issues-Rais

ed-by-Artificial-Intelligence.pdf

35. Can Multi-modal (reasoning) LLMs work as deepfake detectors? - arXiv, accessed May 13, 2025, https://arxiv.org/html/2503.20084
36. Explainable AI for DeepFake Detection - MDPI, accessed May 13, 2025, https://www.mdpi.com/2076-3417/15/2/725
37. (PDF) Explainable AI for DeepFake Detection - ResearchGate, accessed May 13, 2025, https://www.researchgate.net/publication/387994182_Explainable_AI_for_DeepFake_Detection
38. www2.datainnovation.org, accessed May 13, 2025, https://www2.datainnovation.org/2024-ai-watermarking.pdf
39. One Million Euros for Research on Deepfakes in Law Enforcement - Universität Bayreuth, accessed May 13, 2025, https://www.uni-bayreuth.de/en/press-release/research-deepfakes-law-enforcement
40. Developing Robust Solutions, Policies, and Safeguarding Responses to AI-Generated CSAM - ohchr, accessed May 13, 2025, https://www.ohchr.org/sites/default/files/documents/issues/children/sr/cfis/existing-emerging/subm-existing-emerging-sexually-aca-university-toronto-dr-sara-grimes-cewen.pdf
41. EFF Transition Memo to Trump Administration 2025 | Electronic Frontier Foundation, accessed May 13, 2025, https://www.eff.org/wp/eff-transition-memo-incoming-trump-administration
42. Impacts of Adversarial Use of Generative AI on Homeland Security, accessed May 13, 2025, https://www.dhs.gov/sites/default/files/2025-01/25_0110_st_impacts_of_adversarial_generative_al_on_homeland_security_0.pdf