Stanford University
Human-Centered
Artificial Intelligence

Policy Brief
HAI Policy & Society
July 2025

Addressing AI-Generated Child Sexual Abuse Material: Opportunities for Educational Policy

Riana Pfefferkorn

STARTING IN 2023, RESEARCHERS FOUND THAT GENERATIVE AI MODELS were being misused to create sexually explicit images of children. AI-generated child sexual abuse material (CSAM) has become easier to create thanks to the proliferation of generative AI software programs that are commonly called "nudify," "undress," or "face-swapping" apps, which are purpose-built to let unskilled users make pornographic images. Some of those users are children themselves.

In our paper, "AI-Generated Child Sexual Abuse Material: Insights from Educators, Platforms, Law Enforcement, Legislators, and Victims," we assess how several stakeholder groups are thinking about and responding to AI CSAM. Through 52 interviews conducted between mid-2024 and early 2025 and a review of documents from four public school districts, we find that the prevalence of AI CSAM in schools remains unclear but appears to be not overwhelmingly high at present. Schools therefore have a chance to proactively prepare their AI CSAM prevention and response strategies.

Key Takeaways

Most schools are not talking to students about the risks of AI-generated child sexual abuse material (CSAM), specifically via "nudify" apps; nor are they training educators how to respond to incidents of students making and circulating so-called "deepfake nudes" of other students.

While many states have recently criminalized AI CSAM, most fail to address how schools should establish appropriate frameworks for handling child offenders who create or share deepfake nudes.

To ensure schools respond proactively and appropriately, states should update mandated reporting and school discipline policies to clarify whether educators must report deepfake nude incidents, and consider explicitly defining such behavior as cyberbullying.

Criminalization is not a one-size-fits-all solution for minors; state responses to student-on-student AI CSAM incidents should prioritize behavioral interventions over punitive measures, grounded in child development, trauma-informed practices, and educational equity.

The AI CSAM phenomenon is testing the existing legal regimes that govern various affected sectors of society, illuminating some gaps and ambiguities. While legislators in Congress and around the United States have taken action in recent years to address some aspects of the AI CSAM problem, opportunities for further regulation or clarification remain. In particular, there is a need for policymakers at the state level to decide what to do about children who create and disseminate AI CSAM of other children, and, relatedly, to elucidate schools' obligations with respect to such incidents.

Nudify apps' ease of use and lack of restrictions have made them an avenue for the expeditious creation of AI CSAM—including by users who are themselves underage.

The AI CSAM Problem

AI image generation models are abused to create CSAM in several ways. Some AI-generated imagery depicts children who do not exist in real life, though the AI models used to create such material are commonly trained on actual abuse imagery. Another type of AI-generated CSAM involves real, identifiable children, such as known abuse victims from existing CSAM series, famous children (e.g., actors or influencers), or a child known to the person who generated the image. AI tools are used to modify an innocuous image of the child to appear as though the child is engaged in sexually explicit conduct. This type of CSAM is commonly referred to as "morphed" imagery.

The difficulty of making AI CSAM varies. Many mainstream generative AI platforms have committed to combating the abuse of their services for CSAM purposes. Creating bespoke AI-generated imagery depicting a specific child sex abuse scenario thus still entails some amount of technical know-how, such as prompt engineering or fine-tuning open-source models. By contrast, nudify apps,[1] which are trained on datasets of pornographic imagery, take an uploaded photo of a clothed person (either snapped by the perpetrator, or sourced from a social media account, school website, etc.) and quickly return a realistic-looking but fake nude image.

Nudify apps enable those with no particular skills in AI or computer graphics to create so-called "deepfake nudes" or "deepfake porn"—rapidly and potentially of numerous individuals at scale, and typically without the depicted person's consent. What's more, nudify apps do not consistently prohibit the upload of images of underage individuals either in their terms of service or in practice. Their ease of use and lack of restrictions have made them an avenue for the expeditious creation of AI CSAM—including by users who are themselves underage.

Beginning in mid-2023, male students at several U.S. middle and high schools (both public and private) reportedly used AI to make deepfake nudes of their female classmates. A high-profile case in New Jersey was followed by widely reported incidents in Texas,

1 While it can take the form of websites as well as downloadable apps, we refer to all nudify software collectively as "apps" for convenience.

Washington, Florida, Pennsylvania, and multiple incidents in Southern California. More recently, media have reported on additional incidents in Pennsylvania, Florida, and Iowa. There have also been several reported occurrences internationally since 2023.

It is not clear what these cases imply about the prevalence of student-on-student incidents involving deepfake nudes. That they are so spread out geographically could be read to indicate a widespread problem in schools. On the other hand, the number of reported incidents nationwide remains minuscule for a country with over 54 million schoolchildren.

Methodology

In our prior research about online CSAM, we flagged the risk that generative AI might soon unleash a flood of AI CSAM. To follow up on that concern, and in view of the news reports about nudify incidents in schools (which are the focus of this policy brief), we began a new phase of research to investigate AI CSAM's impact on various stakeholders in society including schools.

Between June 2024 and February 2025, we conducted semi-structured interviews with 52 respondents (mostly but not exclusively in the United States), recruited primarily through our existing contacts or by reaching out through publicly available means of contact. Our respondents included nine employees of nongovernmental organizations, eight staff at online platforms (including AI companies), eight groups that provide online trainings for schools, seven academics, six current and former law enforcement officers, four state legislators and state legislative aides, three staff at the National Center for Missing and Exploited Children (NCMEC), two U.S. government employees, two victims (individuals who as children were depicted in images generated with nudify apps), one parent of a victim, one lawyer for a victim, and one teacher. We were unable to secure interviews with school leaders or vendors that offer AI red teaming services.

Using public records requests, we also obtained documents from four public school districts in California, New Jersey, Texas, and Washington about student-on-student deepfake nude incidents at a total of five schools. We identified the schools from news reports about those incidents. We also used news reports, news alerts, and legislative trackers to locate state legislation about AI-generated CSAM that was introduced in 2024 or 2025 as well as state and federal criminal prosecutions begun in 2024 or 2025 that involved AI CSAM or obscene AI-generated depictions of children.

Research Outcomes

We found no consensus among our cross-professional respondents as to the prevalence of nudify incidents in schools. Some had not heard of this being a problem; others believe prevalence is high. We were unable to identify any patterns among those holding either view and are not sure what to make of these conflicting perspectives. The truth may lie somewhere in the middle. Two recent surveys suggest that these incidents are occurring, but if some of our respondents are not hearing much about them, that suggests, at minimum, that relatively few schools have handled a known case. It could also be that schools tend to handle these incidents internally, making prevalence harder to discern.

Why would children make deepfake nudes of their peers? Respondents described a spectrum of intent and motivations. Some may think it is funny; for others, it is a form of bullying; still others engage in malicious at-scale image generation or even sextortion. Multiple respondents emphasized that minors' brains are not yet fully developed, so they cannot fully understand the potential consequences and harmfulness of their actions. Digital interactions can also seem less real or material to children than face-to-face ones. Children might assume that this conduct is legal given nudify apps' availability through app stores, search engines, and social media ads despite those entities' efforts to bar them. However, one recent study found that a large majority of under-18 respondents recognized that making deepfake nude images of other children is illegal.

There is no school-level representative descriptive data on whether students are receiving instruction about online exploitation risks. Several respondents told us that schools often provide instruction on online exploitation only reactively, after an incident. Our interviews suggest that most schools are not currently addressing the risks of nudify apps with students. Whether they *should* is debatable, given the risk that alerting students to such apps' existence might backfire by inducing some students to try using them.

We found that schools often struggle to respond correctly to a student-on-student deepfake nude incident. Teachers generally lack training in how to deal with such incidents, and schools are apt to lack a crisis management plan for this circumstance. Externally, schools sometimes fail to timely report nudify incidents to the appropriate authorities. This may be due to ambiguity in state mandatory-reporter laws' applicability to deepfake nudes, school personnel's insufficient knowledge of state law, or—as some respondents suggested—a desire to sweep the matter under the rug. Internally, schools may be slow to act on a tip or may be more focused on discipline than supporting victims, whose trauma is exacerbated when school staff respond inadequately. When schools mishandle a nudify incident, it destroys the trust of the victims, their families, and other members of the school community.

We found a lack of consensus on the appropriate consequences for children who make and/or circulate deepfake nudes of other children.

We found a lack of consensus on the appropriate consequences for children who make and/or circulate deepfake nudes of other children. Multiple respondents questioned the propriety of criminal charges (which, empirically, appear rare in the United States) given children's above-mentioned cognitive development levels and the range of "criminal blameworthiness" for their behavior. Counseling, school transfer, and education came up as alternatives. There were respondents, however, who rejected the "kids making bad decisions" framing and favored criminal prosecution.

The difference of opinion persisted among the state-level legislators and legislative staffers we interviewed, whom we had contacted for our study because they

had been among the sponsors of the recent raft of state bills that targeted AI CSAM—laws that, with few exceptions, failed to account for children as offenders, not just victims. Asked about the motivation for their bills, legislative respondents described prosecutors stymied from bringing charges for AI CSAM encountered by law enforcement or constituents whose daughters had been victimized by deepfake nudes. Some respondents conceded they did not really have child offenders in mind when crafting their bills, and they were not sure what the appropriate response should be. One legislative respondent characterized such incidents as cyberbullying; another supported criminal punishment for child offenders given what the respondent called the adult nature of the crime.

The national wave of AI CSAM laws do not address schools' obligations with respect to deepfake nude incidents and what to do about the children responsible for them.

Policy Discussion

Dozens of U.S. states have now enacted AI CSAM-related laws, many in just the last two or three years. A number of these new laws target "morphed" imagery of real children, which, while long banned under federal law, had previously gone unaddressed in some states. Deepfake nudes of minors prompted state legislators to start closing that gap—a word that came up repeatedly in our interviews with legislators and legislative staffers, who said their bills enjoyed ample support and little opposition. As one staffer said, "Nobody objects to trying to protect children."

The national wave of AI CSAM laws is a rare example of bipartisan consensus and momentum in a deeply divided era. However, our research identified aspects of the AI CSAM problem that these laws do not address—including, at the state level, schools' obligations with respect to deepfake nude incidents and what to do about the children responsible for them.

First, states should clarify their mandated reporter laws. In most states, school personnel are mandated to report suspected child sexual abuse to law enforcement or child welfare services. "Sexual abuse" includes using a child to produce CSAM, but "nudifying" another child's image does not involve that child's participation, prompting confusion over whether educators must report such incidents; schools that failed to report have faced police inquiries and litigation. States could update their mandatory reporter laws to expressly address deepfake nude incidents as well as the nonconsensual dissemination of a minor's real nude images. Meanwhile, schools and school districts should preemptively set policies about nudify incident reporting and train personnel.

Second, states can induce educators to prepare for student-on-student deepfake nude incidents by defining

them as a form of bullying, for which almost every state requires schools to have policies in place. For example, a proposed Texas bill would have added deepfake nudes to the state's definition of cyberbullying, required school districts to add the production or distribution of deepfake nude images (as well as real nude images) to schools' cyberbullying policies, and allowed public school students to be disciplined for releasing or threatening to release them. Beyond clarifying definitions, policymakers should also encourage schools to put in place detailed plans and allocate resources for responding to student creation of AI CSAM.

Third, and relatedly, child offenders should not be subject to the same consequences as adults who create and/or share deepfake nudes of children. While this behavior is a form of image-based sexual abuse and such images are illegal, we share many of our respondents' deep skepticism about criminalization. The United States already incarcerates youth at higher rates than other countries do. Rather than rush to extend criminal penalties for AI CSAM to juveniles, state policymakers should evaluate various disciplinary options, in consultation with educators (including those skilled in trauma-informed pedagogy), child psychologists, juvenile justice advocates, victims' advocates, and other experts. For example, a few states recently passed AI CSAM laws requiring behavioral health counseling or educational programs for young or first-time offenders.

Looking Ahead

With student-on-student deepfake nude incidents still seemingly limited in scope, now is the time to fill research gaps that could inform smarter school prevention and response strategies. Further research could document how many U.S. schools have experienced incidents involving student-on-student use of nudify apps; whether U.S. schools are training personnel and providing instruction about online exploitation risks to students (and/or to parents and guardians); and the effectiveness of such training. In addition, to build on surveys conducted to date by two organizations, future research could further document the individual-level prevalence among minors in the United States of having been depicted in a "nudified" image.

Now is the time to fill research gaps that could inform smarter school prevention and response strategies.

Student privacy laws limit what research can be conducted in this domain. While essential for protecting young victims, accused students, and witnesses, these laws lead schools to restrict disclosure of details about incidents—such as demographics or context—that could help researchers identify trends or disparities in discipline and support. If schools shared de-identified or aggregated information with policymakers or researchers, that could help shape prevention and response efforts though it might still entail privacy risks. Here, as elsewhere, making use of student data is in tension with protecting student privacy—a perennial issue in education policy.

Reference: The original paper is accessible at Shelby Grossman, Riana Pfefferkorn, and Sunny Liu, "**AI-Generated Child Sexual Abuse Material: Insights from Educators, Platforms, Law Enforcement, Legislators, and Victims**," Stanford Digital Repository, May 29, 2025, https://purl.stanford.edu/mn692xc5736.

Riana Pfefferkorn is a policy fellow at the Stanford Institute for Human-Centered Artificial Intelligence (HAI).

Stanford University's Institute for Human-Centered Artificial Intelligence (HAI) applies rigorous analysis and research to pressing policy questions on artificial intelligence. A pillar of HAI is to inform policymakers, industry leaders, and civil society by disseminating scholarship to a wide audience. HAI is a nonpartisan research institute, representing a range of voices. The views expressed in this policy brief reflect the views of the authors. For further information, please contact **HAI-Policy@stanford.edu**.

Stanford University
Human-Centered
Artificial Intelligence

Stanford HAI: 353 Jane Stanford Way, Stanford CA 94305-5008

T 650.725.4537 **F** 650.123.4567 **E** HAI-Policy@stanford.edu **hai.stanford.edu**