



## **Primer: Drawing the Line Around Innocence in the Generative AI Age**

By: Benjamin Osborne

### **I. Introduction: AI's Rapid Escalation of Harm and the Need for Legal Guardrails**

Artificial intelligence (AI) is transforming how digital content is created, disseminated, and accessed, enabling unprecedented capabilities in generating realistic images, audio, and video that were previously the exclusive domain of human creators. Modern generative AI systems, including large language models (LLMs), diffusion-based image generators, and multimodal reasoning models, are large-scale deep neural networks trained through self-supervised learning on vast datasets of preexisting human-created content collected from the internet and other sources. These neural networks do not store or replay individual works. Instead, through layered parameter adjustments across billions of connections, they learn compressed statistical representations of patterns, structures, and relationships within language and imagery. They then generate novel outputs by sampling from those learned probability distributions, predicting the next token in text, iteratively denoising latent representations in images, or integrating cross-modal signals in multimodal systems. Because their capabilities are derived from patterns extracted across massive datasets of human-created expression, the digital ecosystem itself, including photographs of children such as school portraits, family images, and social media posts, becomes embedded within the training data that shapes and conditions synthetic outputs. The architecture itself creates foreseeable risk.<sup>1</sup>

This generative capacity, while beneficial in many contexts, has dramatically accelerated the creation and spread of harmful content.<sup>2</sup> What can make these matters worse is that this content includes deepfakes, sexually exploitative material involving children, and mechanisms that facilitate grooming and coercive abuse. *AI-generated child sexual abuse material* (AI CSAM) refers to sexually explicit visual depictions of minors produced through generative systems rather

---

<sup>1</sup> Benjamin Osborne, *Primer: A Society in Flux: Copycats, Creativity, and the Future of Artificial Intelligence Governance*, CTR. FOR RENEWING AM. (Dec. 5, 2025), <https://americarenewing.com/issues/primer-a-society-in-flux-copycats-creativity-and-the-future-of-artificial-intelligence-governance/>.

<sup>2</sup> Claudia Cox, *Three Young Men Have Now Taken Their Lives After Disturbing Messages with AI Chatbots*, THE TAB (Sep. 2025), <https://thetab.com/2025/09/03/three-young-men-have-now-taken-their-lives-after-disturbing-messages-with-ai-chatbots>; see also Jeff Horwitz, *Meta's Flirty AI Chatbot Invited a Retiree to New York*, REUTERS (Aug. 14, 2025), <https://www.reuters.com/investigates/special-report/meta-ai-chatbot-death/>.

than through traditional photography involving physical abuse.<sup>3</sup> These depictions may be wholly synthetic or may involve the digital manipulation of an identifiable child’s likeness into explicit imagery. *Deepfakes* are synthetic images, audio, or videos created through machine learning techniques that convincingly depict individuals in events that never occurred.<sup>4</sup> *Grooming* refers to the process by which an adult builds trust and emotional connection with a minor for the purpose of sexual exploitation, often through gradual normalization of sexual content, emotional manipulation, and coercion.<sup>5</sup>

The capacity of modern AI tools to produce AI CSAM poses profound legal, ethical, and enforcement challenges that current law only partially addresses and current Supreme Court precedent only partially covers.<sup>6</sup> Generative systems do not merely reproduce isolated images; they scale exploitation. What once required advanced digital manipulation skills can now be accomplished through simple text prompts and widely available applications. Offenders can use AI tools to alter innocent photographs of minors into explicit depictions, collapsing the distinction between passive possession and active production.

And the harm is tangible. AI-generated CSAM not only replicates exploitative imagery at scale but also normalizes predatory behavior, facilitates grooming through synthetic personas, and enables sextortion through fabricated content. When a child’s likeness is weaponized in this way, the injury radiates beyond the individual victim. Families endure stigma and trauma. Schools confront reputational fallout. Communities lose trust in digital environments. When AI systems transform human expression into raw material for replication, social costs follow.

Federal law already prohibits the creation, possession, distribution, or receipt of CSAM, including computer-generated depictions that are “virtually indistinguishable” from an image of an actual minor.<sup>7</sup> Yet enforcement and statutory clarity remain uneven across jurisdictions. The rise of deepfake and AI tools has outpaced legislative adaptation, leaving gaps that predators can exploit and constitutional doctrine that was crafted in a pregenerative era.

Recent federal policy developments also reflect growing concern about fragmented AI regulation. A December 2025 executive order directed federal agencies to explore a national framework for artificial intelligence and to challenge certain state AI laws that conflict with

---

<sup>3</sup> RAINN, *What About AI-Generated CSAM-Like Deepfakes?* (Aug. 28, 2025), <https://rainn.org/get-the-facts-about-csam-child-sexual-abuse-material/what-about-ai-generated-csam-like-deepfakes/>.

<sup>4</sup> U.S. Gov’t Accountability Off., GAO-20-379SP, *Artificial Intelligence: An Accountability Framework for Federal Agencies and Other Entities* (2020), <https://www.gao.gov/assets/gao-20-379sp.pdf>.

<sup>5</sup> Daniel Pollack & Andrea MacIver, *Understanding Sexual Grooming in Child Abuse Cases*, AMR. BAR ASS’N. (Nov. 1, 2015), [https://www.americanbar.org/groups/public\\_interest/child\\_law/resources/child\\_law\\_practiceonline/child\\_law\\_practice/vol-34/november-2015/understanding-sexual-grooming-in-child-abuse-cases/](https://www.americanbar.org/groups/public_interest/child_law/resources/child_law_practiceonline/child_law_practice/vol-34/november-2015/understanding-sexual-grooming-in-child-abuse-cases/).

<sup>6</sup> See Section III. Existing Legal Framework and Constitutional Constraints, *infra*.

<sup>7</sup> 18 U.S.C. § 2256(8)(B), (11).

federal policy.<sup>8</sup> Notably, the order expressly excludes state laws addressing child safety and exploitation, underscoring the continuing importance of robust protections for minors in the AI ecosystem.<sup>9</sup> A national legal framework for accountability, regulation, and enforcement must therefore balance child protection with constitutional safeguards. It must address likeness-based deepfakes, AI-facilitated grooming, and the structural risks embedded in generative architectures while remaining tailored to survive First Amendment scrutiny. Without such guardrails, the scale and sophistication of synthetic exploitation will continue to grow, and the law will remain reactive rather than protective.

## **II. The Architecture of Harm: Deepfakes, AI CSAM, and Grooming**

### **A. Deepfakes and AI-Generated CSAM**

AI deepfakes are synthetic images or videos produced by machine learning algorithms that can realistically depict individuals in settings or poses that never occurred in reality. Digital forgeries of intimate activity can be created with minimal technical skill and widely distributed online. The Disrupt Explicit Forged Images and Non-Consensual Edits Act (DEFIANCE Act), introduced and repeatedly advanced in Congress, defines digital forgery as a visual depiction created by software, machine learning, or AI that appears realistic to a reasonable observer and is disseminated without consent.<sup>10</sup> This bill, passed unanimously in the Senate in January 2026, extends legal remedies to victims of nonconsensual deepfakes, including civil causes of action for distribution and possession of such imagery.<sup>11</sup>

Even where a deepfake does not feature a real person, it can depict a minor in sexually explicit contexts and cause the same types of harm, including revictimization, grooming, blackmail, and psychological trauma. Because these depictions are “virtually indistinguishable” from authentic imagery, they can functionally serve as CSAM under existing legal enforcement. States like California have already criminalized the creation, distribution, and possession of AI CSAM to close gaps in existing statutory language that predated generative AI technologies.<sup>12</sup>

### **B. Grooming, Coercion, and AI’s Facilitation of Predatory Conduct**

AI can also be misused to facilitate grooming, a process by which abusers cultivate trust and manipulate minors into exploitative relationships. Modern generative models can simulate personas, tailor language to individual psychological cues, and obscure identities, lowering the barrier to predatory conduct that once required significant time and effort. LLM-based systems

---

<sup>8</sup> Ensuring a National Policy Framework for Artificial Intelligence, 90 Fed. Reg. 58499 (Dec. 16, 2025).

<sup>9</sup> *Id.*

<sup>10</sup> *DEFIANCE Act of 2025*, S. 1837, 119th Cong. (2025).

<sup>11</sup> *Id.*

<sup>12</sup> Cal. Penal Code § 311.11.

are designed to predict and replicate patterns of human interaction.<sup>13</sup> When deployed conversationally, this architecture enables AI to mimic emotional bonding, reinforce dependency, and escalate intimacy, which are precisely the behavioral mechanics that define grooming in offline contexts.

AI tools may be programmed or manipulated to initiate and sustain harmful interactions with minors, normalizing explicit content or coercive requests in ways that resemble predatory modi operandi.

### **III. Existing Legal Framework and Constitutional Constraints**

#### **A. CSAM Under Federal Law**

Federal criminal statutes make it unlawful to produce, distribute, receive, or possess CSAM involving minors under the age of eighteen, with penalties that include imprisonment, fines, and registration as a sex offender.<sup>14</sup> The FBI has explicitly warned that AI-generated CSAM is illegal under existing federal law, which covers computer-generated images that appear to depict minors engaging in sexually explicit conduct.<sup>15</sup> Generative AI models trained on massive quantities of open-source text and images use sophisticated machine learning algorithms to learn patterns and relationships from underlying data, enabling even minimally technical users to generate realistic artwork, images, and videos from simple text prompts, including CSAM.<sup>16</sup> What once required advanced digital manipulation skills can now be accomplished through accessible web-based applications.

The distinction between production and possession offenses under federal child exploitation law illustrates how generative AI is reshaping criminal liability. Historically, 18 U.S.C. § 2251 has governed the production of CSAM, targeting those who use or exploit a minor to create new sexually explicit depictions, while 18 U.S.C. § 2252A has focused on the receipt, possession, or distribution of already existing material.<sup>17</sup>

Generative AI collapses this traditional divide. A user who once might have merely downloaded illicit material can now manipulate or transform images into new explicit depictions, potentially triggering the more severe production statute rather than a possession charge. In this sense, AI does not simply expand access to illicit content; it lowers the technical threshold for committing higher-level federal offenses. The law confronts a technological reality in which the line between

---

<sup>13</sup> See Osborne, *supra* note 1.

<sup>14</sup> 18 U.S.C. §§ 2251–2252A.

<sup>15</sup> Fed. Bureau of Investigation, Public Service Announcement, Criminal Actors Continue to Exploit Generative Artificial Intelligence to Produce Child Sexual Abuse Material (Mar. 29, 2024), <https://www.ic3.gov/PSA/2024/PSA240329>.

<sup>16</sup> *Id.*

<sup>17</sup> 18 U.S.C. § 2251; *Id.* § 2252A.

possessor and producer becomes increasingly blurred, escalating both the scale of harm and the severity of criminal exposure.

In November 2023, a Charlotte, North Carolina–based child psychiatrist was sentenced to forty years in federal prison followed by thirty years of supervised release after pleading guilty to sexual exploitation of a minor and using AI to create CSAM.<sup>18</sup> According to federal prosecutors, the individual used AI to alter otherwise innocent images of clothed minors into sexually explicit depictions, thereby producing illegal CSAM under federal law.<sup>19</sup> The case is notable not only for the severity of the sentence but for its confirmation that AI-assisted image manipulation falls squarely within existing federal prohibitions on the production of CSAM under 18 U.S.C. § 2251.<sup>20</sup> What once required direct physical abuse or advanced digital manipulation was instead accomplished through accessible AI tools, demonstrating how generative systems can lower the technical barrier to exploitation while triggering the same criminal liability.

In a separate case that same month, a federal jury convicted a registered sex offender in Pittsburgh, Pennsylvania, for possessing CSAM that had been digitally altered to superimpose the faces of child actors onto nude bodies and bodies engaged in sexual acts.<sup>21</sup> According to the Department of Justice, the defendant possessed modified images in which minors’ likenesses were inserted into sexually explicit scenarios, demonstrating that digital manipulation, whether through traditional editing software or AI-enabled tools, does not shield offenders from criminal liability under federal CSAM statutes.<sup>22</sup> The conviction reinforces the principle that the law focuses not only on the origin of an image but on the exploitation inherent in its creation and possession, confirming that technologically altered depictions of minors in sexual contexts remain prosecutable offenses under existing law.<sup>23</sup>

## **B. CSAM Under State Law**

Nearly every state criminalizes the production, possession, and distribution of CSAM, typically defined as a visual depiction of a minor under eighteen engaged in sexually explicit conduct.<sup>24</sup> These statutes generally mirror federal definitions, including “lascivious exhibition” standards established in *United States v. Dost* and derived from 18 U.S.C. § 2255.<sup>25</sup>

---

<sup>18</sup> U.S. Attorney’s Office for the W. Dist. of N.C., *Charlotte Child Psychiatrist Sentenced to 40 Years in Prison for Sexual Exploitation of a Minor and Using Artificial Intelligence to Create Child Sexual Abuse Material* (Nov. 8, 2023), <https://www.justice.gov/usao-wdnc/pr/charlotte-child-psychiatrist-sentenced-40-years-prison-sexual-exploitation-minor-and>.

<sup>19</sup> *Id.*

<sup>20</sup> *Id.*; 18 U.S.C. § 2251.

<sup>21</sup> U.S. Dep’t of Justice, *Registered Sex Offender Convicted of Possessing Child Sexual Abuse Material* (Nov. 2023), <https://www.justice.gov/archives/opa/pr/registered-sex-offender-convicted-possessing-child-sexual-abuse-material>.

<sup>22</sup> *Id.*

<sup>23</sup> *Id.*; 18 U.S.C. § 2252A.

<sup>24</sup> *See, e.g.*, Cal. Penal Code § 311.11(a); Tex. Penal Code Ann. § 43.26(a); N.Y. Penal Law § 263.16.

<sup>25</sup> 636 F. Supp. 828, 832 (S.D. Cal. 1986); 18 U.S.C. § 2255.

The rise of generative AI has sharpened a key issue: whether digitally altered or synthetic depictions fall within existing statutory language. Some states have explicitly clarified that computer-generated or digitally manipulated images qualify as CSAM.<sup>26</sup> Federal law similarly covers visual depictions that are “indistinguishable” from an actual minor engaged in sexually explicit conduct.<sup>27</sup> But other states rely on older language that presumes the involvement of an identifiable real child, creating interpretive gaps when confronting purely AI-generated deepfakes.

Deepfake exploitation intersects directly with grooming statutes as well. Many states criminalize electronic solicitation, luring, or unlawful contact with a minor through digital communication.<sup>28</sup> AI systems can simulate personas, tailor communications, and facilitate sextortion schemes, effectively accelerating traditional grooming methods. When an offender uses AI to build trust, manipulate a minor, or threaten distribution of fabricated explicit images, existing grooming and exploitation statutes may be triggered before a physical meeting occurs.

### **C. Supreme Court Jurisprudence**

Constitutionally, states retain broad authority to regulate child sexual exploitation. The Supreme Court in *New York v. Ferber* held that child pornography involving real minors is categorically excluded from First Amendment protection.<sup>29</sup> However, there remains a hurdle for an outright ban on all child-related sexual imagery: The Court in *Ashcroft v. Free Speech Coalition* ruled against overly broad bans on purely virtual depictions that do not involve real minors based on constitutional concerns.<sup>30</sup> Legislatures working in the current legal landscape therefore must draft AI-focused statutes with precision, targeting exploitative deepfakes and grooming conduct without sweeping in protected expression.

The result is a patchwork. Most states can prosecute AI-assisted grooming and altered CSAM under existing law. Some have modernized statutes to expressly include synthetic imagery. Others remain vulnerable to interpretive disputes. As generative AI lowers the barrier to creating deepfake abuse and accelerates grooming tactics, these inconsistencies underscore the need for clearer statutory alignment and a coordinated national framework.

#### **1. Overview of *Ashcroft v. Free Speech Coalition***

In *Ashcroft v. Free Speech Coalition*, the Supreme Court struck down portions of the Child Pornography Prevention Act of 1996 (CPPA) that criminalized certain “virtual” depictions of

---

<sup>26</sup> See Cal. Penal Code § 311.11(a).

<sup>27</sup> 18 U.S.C. § 2256(8)(B).

<sup>28</sup> See 18 Pa. Cons. Stat. § 6318(a); N.C. Gen. Stat. § 14-202.3(a).

<sup>29</sup> 458 U.S. 747, 763–64 (1982).

<sup>30</sup> 535 U.S. 234, 256–58 (2002).

minors engaged in sexually explicit conduct.<sup>31</sup> The statute expanded the federal definition of child pornography to include images that “appear to be” of minors, even if no real child was used in producing the material.<sup>32</sup> It also prohibited images that were “advertised, promoted, presented, described, or distributed” in a manner that conveyed the impression that they depicted a minor engaged in sexually explicit conduct.<sup>33</sup>

The Court held those provisions unconstitutional under the First Amendment because they prohibited speech that did not involve the sexual abuse of actual children.<sup>34</sup> Distinguishing *Ferber*, the Court explained that bans on child pornography are constitutional because the production of such material necessarily involved real child abuse.<sup>35</sup> By contrast, purely virtual depictions like computer-generated images or youthful-looking adults did not entail the exploitation of an actual minor and therefore could not be categorically excluded from First Amendment protection.<sup>36</sup>

The Court rejected the government’s argument that virtual child pornography could be banned because it might encourage pedophiles or make prosecution of real CSAM more difficult, emphasizing that “[t]he Government may not suppress lawful speech as the means to suppress unlawful speech.”<sup>37</sup> Accordingly, the Court declared unconstitutional the CPPA’s “appears to be” and “conve[y] the impression” provisions.<sup>38</sup>

After *Ashcroft*, Congress amended the statute’s provisions through the PROTECT Act in 2003, which is why 18 U.S.C. § 2256(8)(b) now reads “virtually indistinguishable” from a minor engaging in sexually explicit conduct rather than the that minors “appear to be” engaging in sexual conduct, language that the Court found unconstitutional in the CPPA.<sup>39</sup>

## 2. How Generative AI Alters the Constitutional Landscape

*Ashcroft* was decided in a technological era that assumed virtual depictions of minors were rare, difficult to produce, and largely disconnected from real-world harm. The Court confronted static, digitally created images that required specialized skill and were unlikely to displace markets for actual CSAM.<sup>40</sup> The underlying premise was clear: Purely virtual depictions did not involve real children and therefore did not implicate the compelling state interest recognized in *Ferber*.<sup>41</sup>

---

<sup>31</sup> *Id.*

<sup>32</sup> *Id.* at 241–42 (quoting 18 U.S.C. § 2256(8)(B) (1996)).

<sup>33</sup> *Id.* at 242 (quoting 18 U.S.C. § 2256(8)(D) (1996)).

<sup>34</sup> *Id.* at 256–58.

<sup>35</sup> 535 U.S. at 249–50.

<sup>36</sup> *Id.* at 250–51.

<sup>37</sup> *Id.* at 255.

<sup>38</sup> *Id.* at 256–58.

<sup>39</sup> 18 U.S.C. § 2256(8)(B), (11).

<sup>40</sup> *Id.* at 250–51.

<sup>41</sup> 458 U.S. at 756–57.

Generative AI collapses those assumptions.

Modern AI systems make production instantaneous and scalable. They dramatically lower technical barriers, allowing minimally skilled users to generate photorealistic depictions of minors with simple text prompts. These systems can fabricate explicit images, insert real children’s faces into sexual content, and create synthetic personas capable of engaging minors in sustained, personalized interaction. Congress has already defined child pornography to include depictions “virtually indistinguishable” from real minors such that an “ordinary person” would believe that the depiction “is of an actual minor engaged in sexually explicit conduct.”<sup>42</sup>

And the harms in *Ashcroft* are no longer abstract. AI-generated deepfakes are used to facilitate grooming through synthetic personas, to carry out sextortion by threatening the dissemination of fabricated images, and to inflict reputational and psychological harm even where no physical abuse occurred. When a child’s likeness is digitally weaponized, the injury is immediate and concrete. The assumption in *Ashcroft* that virtual depictions lack a victim no longer holds in situations involving likeness-based deepfakes or grooming-facilitated exploitation.

Just as *Reno v. ACLU* rested on factual assumptions about the internet’s limited intrusiveness and user-controlled exposure, which are assumptions later eroded by technological development, constitutional doctrine cannot remain frozen in the empirical conditions of a prior technological moment.<sup>43</sup> Where doctrinal premises rest on empirical conditions like the rarity of virtual depictions or the absence of operational real-world harm, those premises warrant reexamination when technology transforms their factual foundation.

#### **D. Level of Scrutiny for Laws Protecting Minors**

Recent Supreme Court jurisprudence suggests an evolving approach to regulations designed to protect minors from harmful online content. In *Free Speech Coalition v. Paxton*, the Court upheld a Texas statute requiring age verification for websites containing a substantial portion of material deemed harmful to minors.<sup>44</sup> Rather than applying strict scrutiny, the Court applied intermediate scrutiny, emphasizing that states retain traditional authority to protect children from sexually explicit material.<sup>45</sup> The Court explained that the First Amendment “leaves undisturbed States’ traditional power to prevent minors from accessing speech that is obscene from their perspective.”<sup>46</sup>

---

<sup>42</sup> 18 U.S.C. § 2256(8)(B), (11).

<sup>43</sup> 521 U.S. 844 (1997); See Adam Candeb, Clare Morell, Joshua Arndt & Hayden Parsons, *Combating Obscenity on the Internet: A Legal and Legislative Path Forward*, CTR. FOR RENEWING AM. (Dec. 15, 2022), <https://americarenewing.com/issues/combating-obscenity-on-the-internet-a-legal-and-legislative-path-forward/>.

<sup>44</sup> 606 U.S. \_\_\_\_ (2025).

<sup>45</sup> Id.

<sup>46</sup> Id.

This reasoning reflects a broader constitutional principle: Where laws are directed at protecting minors from harmful content and only incidentally burden adult access to protected expression, courts may apply a less exacting level of scrutiny. The state’s long-standing role in safeguarding the welfare of children has historically justified regulatory measures that would be impermissible if applied to adults alone. As technological change expands the scale and accessibility of exploitative material, this doctrinal framework provides additional constitutional grounding for narrowly tailored laws addressing emerging forms of digital abuse.

This framework, however, sits in tension with the Court’s earlier decision in *Ashcroft*, which placed constitutional limits on regulating purely virtual depictions of minors. Although the Court treated such material as protected speech, the rapid development of AI raises new questions about whether *Ashcroft*’s reasoning adequately accounts for technologies that can facilitate the types of harms emerging from AI CSAM. The following sections discuss ending *Ashcroft*’s outdated protections.

#### **IV. Generative AI and Reconsidering *Ashcroft***

There can be no constitutional protection for child sexual exploitation, whether the depiction is produced through physical abuse, digital manipulation, or generative artificial intelligence. *Ashcroft* drew a constitutional line between real and virtual child pornography based on the assumption that synthetic depictions were rare and untethered from actual harm. That assumption has collapsed.

Several of the justices saw that danger in 2002. Justice Sandra Day O’Connor warned that “given the rapid pace of advances in computer-graphics technology, the Government’s concern is reasonable,” noting that computer-generated images already “bear a remarkable likeness to actual human beings” and that the Court’s precedents “do not require Congress to wait for harm to occur before it can legislate against it.”<sup>47</sup> That warning has become reality.

Generative AI has erased the practical boundary between fiction and exploitation. Synthetic depictions are scalable, indistinguishable from real children, often derived from actual minors’ images, and routinely deployed in grooming, sextortion, and coercion. The harm is concrete and foreseeable. It does not depend on whether a camera captured physical abuse. It arises from the sexualization, manipulation, and weaponization of children’s likenesses themselves.

*Ashcroft* constitutionalized a technological distinction that no longer exists. The compelling interest recognized in *Ferber* does not evaporate simply because exploitation occurs through pixels rather than photography. When doctrine rests on empirical premises that technology has overtaken, constitutional analysis must adapt. The First Amendment does not compel shelter for synthetic child exploitation in the generative age.

---

<sup>47</sup> *Ashcroft*, 535 U.S. at 264 (O’Connor, J., concurring in the judgment in part and dissenting in part) (citing *Turner Broad. Sys., Inc. v. FCC*, 520 U.S. 180, 212 (1997)).

## A. Reframing Harm in the Generative Era

In *Ferber*, the Court upheld categorical prohibition because child pornography is intrinsically tied to the abuse of children in its production and distribution.<sup>48</sup> The decision rested on the state’s compelling interest in safeguarding the physical and psychological well-being of minors.<sup>49</sup>

AI deepfakes of minors cause real harm even absent physical production abuse. They weaponize a child’s likeness, facilitate coercion and sextortion, permanently damage reputations, normalize deviant sexualization, and function as grooming tools. These harms are direct and foreseeable, no longer just speculative.

The “no victim” logic of *Ashcroft* collapses where a real child’s likeness is digitally manipulated, the synthetic image is used to coerce or groom a minor, or the depiction causes measurable psychological and reputational injury. In these contexts, the state’s interest mirrors that recognized in *Ferber* because the harm is no longer hypothetical. It is operational.

## B. Likeness-Based Deepfakes and Identifiable Minors

*Ashcroft* protected purely fictional depictions untethered to identifiable minors.<sup>50</sup> Generative AI deepfakes, by contrast, frequently involve images of real children’s faces, harvested social media photographs, or depictions of otherwise identifiable minors.

This distinction is constitutionally significant where a real minor’s likeness is digitally inserted into explicit content; the depiction is “virtually indistinguishable” from reality; and the material is tied to grooming, coercion, or reputational injury. This speech is materially different from the abstract, fictional images considered in 2002. The harm flows not from imagination but from the appropriation of a real child’s identity.

## C. Speech Integral to Criminal Conduct

The Supreme Court has long recognized that speech used as an “integral part of [criminal] conduct” is not protected by the First Amendment.<sup>51</sup> When AI-generated explicit images are used as tools of sextortion, blackmail, or grooming, they are not abstract expressions; they are mechanisms of coercion.

States criminalize electronic solicitation and unlawful contact with minors.<sup>52</sup> AI-generated deepfakes deployed in furtherance of such conduct fall squarely within established constitutional exceptions. Even if purely fictional virtual depictions remain protected under *Ashcroft*, synthetic

---

<sup>48</sup> 458 U.S. at 756–58.

<sup>49</sup> *Id.*

<sup>50</sup> 535 U.S. at 256–58.

<sup>51</sup> *Giboney v. Empire Storage & Ice Co.*, 336 U.S. 490, 498 (1949).

<sup>52</sup> *See, e.g.*, 18 Pa. Cons. Stat. § 6318(a); N.C. Gen. Stat. § 14-202.3.

material used to facilitate grooming or extortion is integral to criminal conduct and therefore unprotected.

*Ashcroft* was decided on the premise that virtual depictions were rare, isolated, and disconnected from real-world harm. Generative AI dissolves those premises. Synthetic CSAM is no longer an anomaly; it is a scalable instrument of coercion, grooming, and reputational violence.

Where virtual depictions are used to exploit real minors, manipulate their likenesses, or facilitate criminal conduct, the premise of “no victim” no longer holds. The First Amendment does not require constitutional blindness to technological transformation. When protected abstraction becomes operational exploitation, doctrinal protection narrows accordingly.

## **V. Legal Gaps and Enforcement Challenges**

Current federal and state CSAM laws were drafted in a pre-AI context and often hinge on conceptions of identifiable victims and traditional production mechanisms. Because AI can generate highly realistic depictions without an identifiable photographed child, one might argue that statutory language does not clearly apply to purely synthetic CSAM. Where statutes have not been updated to include “virtually indistinguishable” AI material, enforcement is inconsistent and may fail to capture the full range of exploitative conduct.

Civil remedies depend on a patchwork of state laws that vary widely in scope and enforcement. The absence of uniform federal standards can leave child victims without a predictable avenue for redress. Enforcing existing laws against AI-enabled exploitation is complicated by technical challenges in attributing harmful content to specific actors, tracing dissemination chains, and identifying the precise tools used to generate illicit material. Deepfakes and synthetic content can be shared through decentralized networks, encrypted platforms, and anonymized channels, making law enforcement investigation resource-intensive and technically demanding.

Efforts to regulate AI depiction and distribution must confront First Amendment concerns. Broad prohibitions of generative content risk chilling lawful expression, including documentary, artistic, and educational uses. Legal frameworks must therefore use narrowly tailored definitions, focusing on sexual exploitation and abuse, as this category of speech falls outside First Amendment protection.<sup>53</sup> Drafting statutes to define prohibited deepfakes in terms of explicitness, lack of consent, and the presence of minors helps ensure that regulation targets harmful content without sweeping in protected speech.

## **VI. Recommendations: A National Framework for AI Accountability and Child Protection**

The rapid evolution of generative AI demands more than piecemeal reform. A coherent national framework must clarify statutory boundaries, reinforce constitutional guardrails, impose

---

<sup>53</sup> 458 U.S. at 763 (1982).

structural safeguards on AI developers and deployers, and provide meaningful remedies to victims. The goal is not to suppress lawful expression but to ensure that synthetic technologies do not become scalable instruments of child exploitation. The following recommendations seek to narrow constitutional ambiguity while strengthening child protection.

### **A. Statutory Clarification and Constitutional Precision**

Congress can amend federal child exploitation statutes to expressly clarify that

- AI-generated depictions that are “virtually indistinguishable” from real minors engaged in sexually explicit conduct constitute CSAM under 18 U.S.C. § 2256(8)(B), even where no physical abuse occurred in production
- the digital manipulation of an identifiable minor’s likeness into sexually explicit content constitutes “production” of exploitative material under 18 U.S.C. § 2251
- the use of AI-generated imagery as leverage in grooming, sextortion, coercion, or blackmail constitutes aggravated exploitation

Such clarification can seek to overrule *Ashcroft* as no forms of child pornography can be constitutionally protected. Where a real child’s likeness is appropriated, where synthetic imagery causes measurable psychological harm, or where such content is integral to grooming or extortion, the “no victim” assumption collapses. Legislative reform can codify this distinction clearly and narrowly.

### **B. Codification of a Likeness-Based Deepfake Standard**

To survive First Amendment scrutiny, statutory reform could distinguish between purely fictional synthetic depictions with no identifiable minor and deepfakes involving the likeness, image, or identifiable characteristics of a real child.

Federal law can define unlawful deepfakes involving minors as synthetic depictions that insert, manipulate, or replicate the likeness of an identifiable minor into sexually explicit content or that are used to facilitate grooming, coercion, or extortion.

As part of overturning the “appears to be” language that the Court in *Ashcroft* called “unconstitutional,” federal law can specify that any sexual depiction of a minor that “appears to be” an actual minor is not protected by the First Amendment.

### **C. Adjustments to Federal Law**

Federal action on deepfakes includes the Take It Down Act, now law, which criminalizes the posting of nonconsensual intimate imagery, including deepfakes, and requires platforms to

remove such content within forty-eight hours of notice.<sup>54</sup> This legislation, while oriented toward intimate imagery broadly, represents a foundation that could be strengthened to address child-centered harms more directly. The DEFIANCE Act, passed in the Senate with bipartisan support, would allow victims to pursue federal civil claims against those who create, distribute, or possess nonconsensual deepfakes, creating a private right of action with liquidated damages, injunctive relief, and privacy protections for plaintiffs.<sup>55</sup>

## 1. Build on the Take It Down Act

The Take It Down Act establishes criminal penalties and rapid removal requirements for nonconsensual intimate imagery, including deepfakes.<sup>56</sup>

Its framework could be expanded to

- include explicit child-centered provisions
- mandate platform compliance audits
- require preservation of evidence for law enforcement
- impose escalating penalties for noncompliance

Rapid removal is critical where minors are involved, as digital permanence magnifies harm.

## 2. Enact and Expand the DEFIANCE Act

The DEFIANCE Act provides a federal civil cause of action for victims of nonconsensual deepfakes.<sup>57</sup>

Congress can

- Create a federal civil remedy for an identifiable individual whose intimate image has been digitally forged without consent
- authorize suits against a person who knowingly produces, discloses, distributes, solicits, or possesses such content with intent to distribute a nonconsensual digital forgery
- define *digital forgery* as an intimate visual depiction created or altered using software, machine learning, or artificial intelligence that is indistinguishable to a reasonable person from an authentic image of the identifiable individual
- clarify that labeling an image as fake does not eliminate liability
- provide a ten-year statute of limitations measured from the discovery of the violation or from when the victim turns eighteen, whichever is later

---

<sup>54</sup> *Take It Down Act*, Pub. L. No. 119-12 (2025).

<sup>55</sup> *DEFIANCE Act of 2025*, S. 1837, 119th Cong. (2025).

<sup>56</sup> *Take It Down Act*, Pub. L. No. 119-12 (2025).

<sup>57</sup> *DEFIANCE Act of 2025*, S. 1837, 119th Cong. (2025).

- authorize courts to protect victims’ privacy through pseudonymous litigation and sealing of sensitive materials
- preserve state and tribal laws that provide equal or greater protections
- include a severability clause to safeguard the act against partial constitutional invalidation

### **3. Codify the “Integral to Criminal Conduct” Standard for AI Grooming**

Federal law can clarify that AI-generated content used as an instrument of grooming, coercion, or sextortion falls within the well-established exception for speech integral to criminal conduct.

Where synthetic imagery is deployed

- to threaten dissemination,
- to induce sexual acts,
- to coerce compliance, or
- to facilitate exploitation,

it can be treated not as protected expression but as a tool of criminal conduct. This statutory clarification reinforces existing doctrine without expanding unprotected categories beyond established precedent.

## **C. Developers and Platform Accountability**

### **1. Duties of Care**

Generative systems rely on massive datasets composed of existing human-created content. When minors’ images are embedded within that data ecosystem, misuse becomes structurally foreseeable.

Congress can establish strict civil liability for the commercial development, training, or deployment of generative AI systems that produce CSAM or likeness-based sexually explicit depictions of minors, regardless of intent or knowledge. Liability would attach upon proof that the system generated unlawful material.

This framework operates analogously to product liability: When a commercial actor introduces into commerce a system capable of generating unprotected exploitative material involving minors and that material is produced, responsibility follows from deployment itself. Where exploitation in the CSAM context is a foreseeable byproduct of system design, responsibility cannot be confined to downstream users.

This approach is constitutionally sound because it targets categories of speech already unprotected under *Ferber* or falling within the “speech integral to criminal conduct” doctrine.<sup>58</sup>

---

<sup>58</sup> Giboney, 336 U.S. at 498.

The statute would regulate only CSAM as defined in 18 U.S.C. § 2256(8)(B) and likeness-based exploitative depictions of minors.

Unlike a negligence regime, by placing responsibility on those who design and deploy these systems, strict liability ensures that the burden of precaution rests with the actors best positioned to prevent harm. Developers would be compelled to internalize structural risks and implement safeguards like age-sensitive filters and protective measures against grooming and CSAM usage.

## **2. Obligations**

Because generative systems and LLMs replicate patterns drawn from massive datasets of existing human-created content, responsibility cannot be confined to downstream user misconduct. Where the underlying data ecosystem includes images of minors, exploitative recombination is a predictable byproduct of design choices. Regulatory obligations must therefore address architectural safeguards, not merely user intent.

Developers of AI models and platform providers could be subject to statutory duty of care standards requiring

- age-sensitive filters and safeguards that prevent generation of exploitative content involving minors
- mandatory reporting mechanisms for suspected AI-generated CSAM with law enforcement integration
- retention of logs and audit trails to assist investigations and provide accountability
- transparency and explainability standards for model decisions that materially affect safety

Platforms that host user interactions involving AI generation can be required to implement robust content moderation, including automated detection tools calibrated to identify exploitative patterns and mechanisms to expedite removal.

## **VII. Conclusion**

Generative artificial intelligence has altered the terrain of child exploitation in ways the law did not anticipate. What once required physical proximity, technical sophistication, or clandestine networks can now be accomplished through scalable, automated systems trained on the accumulated digital record of human life. When children's images become raw material for synthetic recombination, exploitation is no longer incidental but becomes structurally foreseeable. The constitutional framework developed in *Ferber* and refined in *Ashcroft* was crafted for a technological era in which virtual depictions were rare, difficult to produce, and largely detached from real-world harm. That era has ended. Synthetic sexualization of minors

---

now facilitates grooming, coercion, reputational destruction, and long-term psychological injury, even absent physical abuse in production.

The escalation of harms to children enabled by generative AI demands a national legal framework that balances enforcement rigor with constitutional safeguards. Children's safety, psychological well-being, and dignity must remain at the forefront of AI governance. Building on existing statutes prohibiting CSAM and recent federal efforts with the Take It Down Act and the civil remedies introduced through the DEFIANCE Act, Congress and the executive branch can pursue statutory clarification, civil accountability, development obligations, and coordinated enforcement to ensure that AI technology does not become a scalable tool of exploitation.

Absent these reforms, AI's power will continue to be wielded without adequate guardrails, leaving children vulnerable to harms that are both foreseeable and preventable. The law must respond with precision, not panic; with constitutional discipline, not overbreadth; but also with moral clarity. A national framework that overturns *Ashcroft*, strengthens protections against likeness-based deepfakes, imposes structural safeguards on generative systems, and empowers families with meaningful remedies is not an abandonment of free expression but a reaffirmation that the First Amendment does not require indifference to technological evolution when that evolution sacrifices children's innocence. The measure of a free society is not merely how it protects speech but *how it protects its children*. In the generative age, drawing that line has never been more imperative.