Addressing Computer-Generated Child Sex Abuse Imagery: Legal Framework and Policy Implications

Riana Pfefferkorn*

February 2024

Computer-generated child sex abuse imagery poses significant challenges to law enforcement, including constitutional limits on criminal prosecutions.

The year 2023 saw significant technical advances in generative machine learning (ML), particularly in the field of image synthesis. Using pictures or text-based prompts as input, content creators can generate new images using "diffusion models," a form of visual generative ML model trained on a data set of images.

When trained on sexually explicit imagery, diffusion models can be used to generate realistic-looking new explicit content. This is now easily accomplished on consumer-grade hardware thanks to rapid advances in visual generative ML models.[1] With augmentations and updated models, image creators can

Riana Pfefferkorn is a research scholar at the Stanford Internet Observatory (SIO). Her Mastodon handle is @riana@mastodon.lawprofs.org. Thank you to David Thiel, SIO's chief technologist, for generously assisting with this paper, especially the background on ML diffusion model technology and information about his paper with Thorn (cited below).

[1] David Thiel, Melissa Stroebel, & Rebecca Portnoff, "Generative ML and CSAM: Implications and Mitigations," Thorn & the Stanford Internet Observatory, June 2023, at 4–7, https://perma.cc/65KW-6ELV (hereafter SIO/Thorn Paper). Diffusion models were initially released with few safety controls in place to prevent training them on explicit content, and versions of those early models remain in use notwithstanding

make rapid and fine-grained adjustments to image content, as well as the age, identity, and pose of human subjects. Customized models for portraying specific sex acts have also proliferated on public sites. Consequently, the open-source ML community now creates explicit content that is routinely mistaken for real.

This usage is not limited to explicit content depicting adults. Starting in early 2023, child-safety researchers began detecting the use of visual generative ML models to create highly realistic child sex abuse material (CSAM). Such images account for a small but growing percentage of the material shared in underground CSAM-trading communities online.[2] There is some evidence that these communities have retrained diffusion models, sometimes on photos of existing CSAM victims.[3] Given the current pace of technological advancement in the field of generative ML, it will soon become significantly easier to generate images that are indistinguishable from actual photographic images depicting the sexual abuse of real children.

While there are some mitigations that can help detect and limit the spread of computer-generated CSAM made with generative ML (CG-CSAM),[4] such images are likely to continue to proliferate for some time. Their rising prevalence will have major implications for the current U.S. legal regime governing child sex abuse imagery. In this paper, I provide an overview of the governing law, discuss the constitutional and policy implications of CG-CSAM, and suggest some potential responses.

With respect to the First Amendment, abuse material produced using real children is constitutionally unprotected. Specifically, CG-CSAM may be criminalized if either it depicts an actual, identifiable child (a "morphed image," for short) or its training data set included actual abuse imagery. Otherwise, the image is likely First Amendment–protected speech under existing Supreme Court precedent concerning

---

the release of later official versions with additional controls. Id. at 2. Consequently, while the corporate terms of use for ML models often prohibit pornographic or illegal uses, those terms cannot be thoroughly enforced in practice. Internet Watch Foundation, "How AI Is Being Abused to Create Child Sexual Abuse Imagery," October 2023, at 15–16, https://perma.cc/2NAH-5AAC (hereafter IWF Report).

[2] SIO/Thorn Paper, supra note 1, at 2–3; IWF Report, supra note 1, at 7–8.

[3] Augmentation of standard-issue diffusion models can be accomplished using low-rank adaptations (LoRAs), small files that "fine-tune" the model with regard to a specific subject (such as a fictional character or real person), pose or act, or artistic style. Since LoRAs for diffusion models are trained on image files, they can be used to fine-tune diffusion models to generate new imagery of existing CSAM victims. SIO/Thorn Paper, supra note 1, at 5–6; Yubin Ma, "The Ultimate Stable Diffusion LoRA Guide (Downloading, Usage, Training)" (version dated Aug. 22, 2023), https://perma.cc/L5BT-UYLX /.

[4] See SIO/Thorn Paper, supra note 1, at 9–16. That paper uses "CG-CSAM," not "AI CSAM" or "ML CSAM," to refer to images created using generative ML; this paper retains that terminology.

"virtual child pornography" unless it qualifies as obscenity, which is also unprotected speech (albeit for different reasons than "real" CSAM).

Although prosecutors and courts must account for the constitutionally significant distinctions among different types of CG-CSAM, online platforms, which are required by law to report CSAM on their services, likely won't. Rather, they will report both CG-CSAM and CSAM depicting real children and let downstream authorities sort it out, adding to a reporting pipeline already overburdened by a high volume of reports. Congress has an opportunity to ameliorate CG-CSAM's effects by investing in technologies for authenticating image provenance and by carefully crafting new legislation narrowly targeting ML models that were trained on actual, photographic abuse imagery. But it should resist the easy temptation to propose unconstitutional responses to the emergent CG-CSAM problem.

EVOLUTION OF CURRENT FEDERAL CSAM LAW

Federal law prohibits both "child pornography" (now usually called "CSAM")[5] and obscene visual depictions involving minors, meaning any person under age 18. Both of these prohibitions will assume new relevance in an age of realistic artificial intelligence (AI)-generated CSAM. Before delving into the statutes, it is crucial to understand the distinction between CSAM and obscenity. The First Amendment does not protect either type of speech, but for different reasons.

Obscenity lacks First Amendment protection because it is devoid of "even the slightest redeeming social importance," and "any benefit that may be derived from [it] is clearly outweighed by the social interest in order and morality."[6] Whether particular material qualifies as legally obscene is determined by the Supreme Court's three-part test in *Miller v. California* (1973). To be obscene, a work must (1) taken as a whole, appeal to the prurient interest, as adjudged by the average person applying contemporary community standards; (2) depict or describe, in a patently offensive way, sexual conduct (as specifically defined by applicable state law); and (3) taken as a whole, lack serious literary, artistic, political, or scientific value.[7]

Nine years after *Miller*, the Supreme Court held in *New York v. Ferber* (1982) that pornography produced using real children is unprotected speech, an exclusion justified by the state's compelling interest in safeguarding children from the myriad harms caused by sexually exploiting them in pornographic materials. That interest is not reflected in the *Miller* test, the *Ferber* Court explained, so obscenity law does not suffice to prohibit CSAM. The harms caused to the children exploited in CSAM

[5] The term "child pornography" is now disfavored, so wherever possible this paper uses "child sex abuse material" or "CSAM" instead. It also uses "photographic CSAM" rather than "real CSAM" or "true CSAM" to refer to imagery of real children, because computer-generated abuse imagery can still do real harm.

[6] *Roth v. United States*, 354 U.S. 476, 484–85 (1957) (quotation omitted).

[7] *Miller v. California*, 413 U.S. 15, 24 (1973).

are unconnected to the *Miller* inquiries of whether the CSAM, taken as a whole, appeals to the prurient interest or has serious value.[8]

In short, not all obscenity is CSAM; obscenity is unprotected speech whether it involves children or not. And, less intuitively, not all CSAM is legally obscene. (Consider, for example, a journalist's photo recording a soldier's sexual abuse of a child in a war zone.) To be sure, many if not most sexually explicit works depicting children will fall into both categories. But CSAM and obscenity are distinct concepts with distinct rationales for placing them outside the First Amendment.

Hence their prohibition by distinct parts of federal criminal law. The relevant statutes are codified in Title 18 of the United States Code, chapters 71 (obscenity) and 110 (sexual exploitation and other abuse of children). Most pertinent for purposes of this paper (because they are the most readily applicable to computer-generated CSAM) are two particular laws: 18 U.S.C. § 2252A, a child pornography statute, and 18 U.S.C. § 1466A, an obscenity statute.

Section 2252A: Child Pornography

First enacted in 1996, Section 2252A prohibits the knowing possession, receipt, or dissemination of "child pornography," as defined in 18 U.S.C. § 2256(8).[9] That definition has changed over time. The 1996 law covered virtual child pornography, out of concern that new technologies allowed "computer imaging … to create [sexually explicit] realistic images of children who do not exist."[10] After the Supreme Court struck down that portion of the definition in 2002, Congress revised the language to its current wording, which continues to cover certain computer-generated images.

Ashcroft v. Free Speech Coalition

Congress's 1996 legislation defined "child pornography," in relevant part, to include "any visual depiction, including any photograph, film, video, picture, or computer or computer-generated image or picture," that "is, or *appears to be*, of a minor engaging in sexually explicit conduct."[11]

---

[8] *New York v. Ferber*, 458 U.S. 747, 760–01 (1982) (finding the *Miller* standard unsatisfactory as a "solution to the child pornography problem" since it "does not reflect the State's particular and more compelling interest in prosecuting those who promote the sexual exploitation of children" and its factors "bear[] no connection to the issue of whether a child has been physically or psychologically harmed in the production of the work").

[9] Pub. L. 104–208, 110 Stat. 3009, 3009–26, 3009–28 (1996).

[10] *Ashcroft v. Free Speech Coalition*, 535 U.S. 234, 240, 241 (2002).

[11] 18 U.S.C. § 2256(8)(B) (1996) (emphasis added).

This definition soon drew a constitutional challenge because it prohibited "speech that is neither obscene under *Miller* nor child pornography under *Ferber*."[12] In 2002's *Ashcroft v. Free Speech Coalition*, the Court reaffirmed its prior holdings that the First Amendment permits the banning of pornography only so long as it is obscene or involves actual minors, meaning it protects "virtual CSAM" that does not involve harm to any actual children.

Justice Anthony Kennedy, writing for a five-justice majority, struck down the challenged portion of the "child pornography" definition as substantially overbroad. Legislation is unconstitutionally overbroad if, in banning unprotected speech (such as photographic CSAM), "a substantial amount of protected speech"—that is, virtual CSAM—"is prohibited or chilled in the process." Congress may ban CSAM's production, distribution, sale, and possession, the Court explained, because those acts harm children. CSAM is "the product of child sexual abuse," and its distribution and sale are "intrinsically related" to that abuse because image circulation harms the abused child and the traffic in CSAM creates an economic incentive to produce it. And banning possession targets the ultimate "level[] in the distribution chain." By contrast, the Court found, virtual CSAM "do[es] not involve, let alone harm, any children in the production process"; it is "not intrinsically related to the sexual abuse of children," "records no crime[,] and creates no victims by its production."[13]

Since the "direct harm" rationale for prohibiting photographic CSAM did not apply to virtual CSAM, the government raised various other arguments to justify the virtual CSAM ban, all of which failed.

- *Indirect harms:* The government argued that *virtual* CSAM could indirectly *lead* to actual abuse. The Court responded that neither virtual CSAM's potential to inspire a pedophile to commit abuse, nor its potential use by a pedophile to seduce children, justifies banning it. The Court had long held that the state may not prohibit speech just because it might tend to encourage unlawful acts or is unfit for children to view.

- *Market deterrence:* The government also argued that its goal of eliminating the market for photographic CSAM necessitated a ban on virtual CSAM, which it contended was part of the same market. The Court rejected that argument: With photographic CSAM, "the creation of the speech is itself the crime of child abuse" that can be deterred by prohibiting the speech, thereby "removing the profit motive" to engage in the crime; with virtual CSAM, "there is no underlying crime" to deter since no real children are harmed, so a "market deterrence" theory could not justify prohibiting such speech.

- *Prosecutorial necessity:* The government argued that virtual CSAM must be banned because it complicates prosecutions. Congress had found that improvements in imaging technology make it "more difficult to prove that a particular picture was produced using actual children," so it

[12] *Free Speech Coalition*, 535 U.S. at 240.

[13] Id. at 241, 249–250, 255–56 (cleaned up).

> banned virtual CSAM to keep the producers and possessors of actual, photographic CSAM from evading prosecution. The Court retorted that it "turns the First Amendment upside down" to argue that "protected speech may be banned as a means to ban unprotected speech": To the contrary, the overbreadth doctrine prohibits the state from sweeping in substantial amounts of protected speech as part of a ban on unprotected speech.

For these (and other) reasons, the Court struck down the "appears to be" portion of the definition of "child pornography."[14]

The PROTECT Act of 2003

In response to *Free Speech Coalition*, Congress passed the PROTECT Act of 2003, which revised the definition of "child pornography."[15] Recall that the definition invalidated by the Court was "[a] computer or computer-generated image or picture" that "is, or appears to be, of a minor engaging in sexually explicit conduct."[16] The PROTECT Act replaced that language with "a digital image, computer image, or computer-generated image that is, or is indistinguishable from, that of a minor engaging in sexually explicit conduct," where "indistinguishable" means "virtually indistinguishable, in that the depiction is such that an ordinary person viewing the depiction would conclude that the depiction is of an actual minor engaged in sexually explicit conduct."[17]

It is hard to see how the new language ("indistinguishable from") differs meaningfully from the old ("appears to be"). Indeed, multiple commentators have argued that the updated language still violates the First Amendment because it continues to proscribe virtual CSAM.[18] This was recognized at the time of the PROTECT Act's enactment, when a group of senators predicted that the definition "will be the

---

[14] Id. at 250–55.

[15] Pub. L. 108–21, 117 Stat. 650, 678–82 (Apr. 30, 2003).

[16] *Free Speech Coalition*, 535 U.S. at 241 (quoting former 18 U.S.C. § 2256(8)(B)).

[17] 18 U.S.C. § 2256(8)(B), (11) (excluding "drawings, cartoons, sculptures, or paintings" from "indistinguishable" definition).

[18] Brian G. Slocum, "Virtual Child Pornography: Does It Mean the End of the Child Pornography Exception to the First Amendment?," *Albany Law Journal of Science & Technology* 14 (2004): 637; Rosalind E. Bell, Note, "Reconciling the PROTECT Act With the First Amendment," *New York University Law Review* 87 (2012): 1877; Congressional Research Service, "Child Pornography Produced Without an Actual Child: Constitutionality of 108th Congress Legislation," Mar. 31, 2003, at 15–16 (analyzing the PROTECT Act before its enactment).

center of much constitutional debate" and warned that "these new definitional provisions risk crossing the constitutional line."[19]

Yet this has remained a purely theoretical argument: For whatever reason, there have been no court opinions addressing the constitutionality of the revised language.[20] This may be simply because virtual CSAM historically was not very realistic and thus did not meet the PROTECT Act's indistinguishability requirement.[21] Rather, that clause was "best interpreted as a forward-looking provision, to be applied, if at all, if and when technology exists to create images that are 'indistinguishable from' actual minors."[22]

That day may have finally come: If diffusion models now enable the creation of truly photorealistic CG-CSAM, courts may soon have to rule on the constitutionality of banning CG material that is "indistinguishable from" photographic CSAM. To date, it remains the law of the land that virtual CSAM is protected speech. In 2008's *United States v. Williams*, which involved the PROTECT Act (but not the "indistinguishable from" language), the Supreme Court affirmed the legality of the market for "simulated" child pornography, "so long as it is offered and sought as such, and not as real child pornography." The Court repeated its reasoning from *Free Speech Coalition* that "the child-protection rationale for speech restriction does not apply to materials produced without children."[23] The Court has not revisited the constitutionality of virtual CSAM since *Williams*.

Section 1466A: Child Obscenity

Along with revising the definition of "child pornography," the PROTECT Act also added a new "child obscenity" statute, codified at 18 U.S.C. § 1466A. Section 1466A prohibits (in specified circumstances) the knowing production, receipt, distribution (or possession with intent to distribute), or possession of the "visual depiction" of a minor that both contains sexually explicit conduct (per the same definition as for Section 2252A) and is obscene. The statute covers "any kind" of "visual depiction," defined to

[19] S. Rep. No. 108-2 (2003) (additional views of Senators Leahy, Biden, and Feingold), https://perma.cc/846C-8RA4.

[20] One defendant did try to challenge the definition under *Free Speech Coalition* for prohibiting virtual CSAM, but the court rejected the argument (without evaluating its merits) because that definition applies only to Section 2252A whereas the defendant had been indicted under a different statute. *United States v. Sherr*, 400 F. Supp. 2d 843, 848–49 (D. Md. 2005).

[21] *Williams*, 553 U.S. at 290–91.

[22] *United States v. Dean*, 670 F. Supp. 2d 1285, 1289 (M.D. Ala. 2009).

[23] *United States v. Williams*, 553 U.S. 285, 289, 303 (2008) (citing *Ashcroft v. Free Speech Coalition*, 535 U.S. 234, 249–51, 254 (2002)).

include any "digital image or picture, computer image or picture, or computer generated image or picture." The statute expressly does not require "that the minor depicted actually exist."[24]

In the PROTECT Act's findings, Congress, citing *Miller* and *Ferber*, recognized the distinct rationales for prohibiting obscenity and CSAM and expressed great dismay at the *Free Speech Coalition* decision.[25] The PROTECT Act thus may be understood as a "belt and suspenders" approach to protecting children post-*Free Speech Coalition*: Even if the Supreme Court were to strike down the act's new definition of virtual CSAM like it had the previous one, Section 1466A would stand on the separate and firmer ground of obscenity doctrine, under which, as noted above, the distinction between real and virtual children is irrelevant.

Accordingly, the crux of Section 1466A is not whether an actual child was used in producing the accused images but, rather, whether they meet the *Miller* obscenity test. Subsections (a)(1) and (b)(1) require that the visual depiction be "obscene." This language has been repeatedly upheld on the grounds that it incorporates *Miller*.[26] In affirming that *Miller* provides the proper test, courts have refused to extend *Free Speech Coalition*'s "no harm to an actual child" reasoning from 2252A to 1466A: "The fact that this statute does not require that an actual minor exist … is immaterial because, unlike" 2252A, 1466A "is an obscenity statute and not a child pornography statute."[27] In short, "[w]hile [*Free Speech Coalition*] did recognize that, in some cases, virtual images are protected by the First Amendment, the case did not extend that protection to obscene virtual images."[28]

It bears noting that Section 1466A has rarely been used. Since the PROTECT Act's passage 20 years ago, there have been over fifty-fold more federal court decisions citing 2252A than 1466A.[29] The child obscenity statute has simply never been prosecutors' preferred vehicle for combating child exploitation. The reason for this, while unclear, might lie in the relative difficulty of prosecution: The knowing possession, receipt, or distribution of (photographic) CSAM is tantamount to a strict liability offense,

---

[24] U.S.C. § 1466A(c), (f)(1).

[25] 117 Stat. at 676–78.

[26] *United States v. Buie*, 946 F.3d 443, 445–46 (8th Cir. 2019); *United States v. Dean*, 635 F.3d 1200, 1205 & n. 4 (11th Cir. 2011); *United States v. Schales*, 546 F.3d 965, 970–72 (9th Cir. 2008); *United States v. Handley*, 564 F. Supp. 2d 996, 1007 (S.D. Iowa 2008). Subsections (a)(2) and (b)(2), which controversially employ a lower bar for liability than the *Miller* test, are almost never used.

[27] *Schales*, 546 F.3d at 971–72 (case involving morphed images); *accord United States v. Whorley*, 550 F.3d 326, 330, 335–37 (4th Cir. 2008) (case involving "obscene Japanese anime cartoons").

[28] *United States v. Mees*, No. 09-cr-145, 2009 U.S. Dist. LEXIS 48801, at *11 (E.D. Mo. June 10, 2009) (citing *Free Speech Coalition*, 535 U.S. at 246).

[29] The results of LexisNexis Shepardizing I conducted on Dec. 20, 2023, showed 7,475 cases citing 2252A and only 138 citing 1466A.

whereas an obscenity case entails the more probing inquiry of the three-pronged *Miller* test. As this paper will discuss, however, the rise of photorealistic CG-CSAM may prompt prosecutors to rediscover Section 1466A if proving obscenity becomes easier than proving the involvement of a real child.

HOW WILL THE FIRST AMENDMENT APPLY TO CG-CSAM?

CG-CSAM's constitutionality is likely to become an issue for the courts. This section analyzes whether CG-CSAM may constitutionally be prohibited under Section 2252A and Section 1466A. Per *Free Speech Coalition*, "virtual child pornography" that does not depict real children is protected speech. Section 2252A's prohibition of computer-generated material that is "indistinguishable" from photographic CSAM has yet to be challenged, but that might change as photorealistic CG-CSAM becomes more common. By contrast, whether material is real or computer-generated is irrelevant to Section 1466A, an obscenity statute whose constitutionality has largely been upheld. If a piece of CG-CSAM was not generated using any actual abuse imagery, does not depict a real child, and is not legally obscene, then it is likely to be deemed protected speech. Otherwise, it may constitutionally be criminalized under existing laws.

Constitutionality of CG-CSAM Under Section 2252A

Section 2252A generally prohibits knowingly possessing, receiving, or disseminating "child pornography."[30] As currently defined, "child pornography" includes "any visual depiction" that (1) "involves the use of a minor engaging in sexually explicit conduct," (2) "is a digital image, computer image, or computer-generated image that is, or is indistinguishable from, that of a minor engaging in sexually explicit conduct," or (3) "has been created, adapted, or modified to appear that an identifiable minor is engaging in sexually explicit conduct."[31]

In this paper, I use "photographic CSAM" (or "actual abuse imagery") to refer to images of actual minors engaging in sexually explicit conduct—that is, CSAM involving real children. I refer to "computer-generated" child sex abuse images as "virtual CSAM" (a term borrowed from *Free Speech Coalition*)—that is, they are not photographs of real children. What this paper calls CG-CSAM—images created using generative ML specifically—is a type of virtual CSAM. Finally, I refer to "adapted or

[30] 18 U.S.C. § 2252A.

[31] Id. § 2256(8). "Sexually explicit conduct" has its own lengthy definition. Id. § 2256(2). An "identifiable minor" means someone "whose image as a minor was used in creating, adapting, or modifying the visual depiction" and "who is recognizable as an actual person." Id. § 2256(9).

modified" images of identifiable minors as morphed images, so called "because they morph a non-sexual image of an identifiable child with sexually explicit images."[32]

Applied to CG-CSAM (i.e., virtual CSAM made using generative ML), Section 2252A appears to prohibit:

- *Abuse-trained CG-CSAM:* Knowing possession (or receipt, distribution, etc.) of CG-CSAM that, while it is virtual, nevertheless "involves the use of" actual abuse because it was generated using training data that included photographic CSAM, whether or not the resulting image "is indistinguishable from" photographic CSAM.
- *Photorealistic CG-CSAM:* Knowing possession (or receipt, distribution, etc.) of ML-generated virtual CSAM that "is indistinguishable from" photographic CSAM, whether or not the model's training data included photographic CSAM.
- *Morphed-image CG-CSAM:* Knowing distribution or production with intent to distribute of ML-generated morphed images depicting an identifiable child, whether or not the model's training data included photographic CSAM (of the identifiable child or other children), and whether or not the morphed image is photorealistic (that is, indistinguishable from photographic CSAM).

These categories can overlap[33]—a photorealistic image or a morphed image may be abuse-trained; a morphed image may be photorealistic—but need not overlap for liability to attach; CG-CSAM that falls into any one of the three categories is covered by Section 2252A. These categories consider both the *input* and *output* of the ML model used to generate the CG-CSAM in question. On the input end, did the model's training data include photographic CSAM? On the output end, is the generated image photorealistic, or does it depict an identifiable minor?

If CG-CSAM nominally violates Section 2252A, that does not end the inquiry. As *Free Speech Coalition* demonstrates, the First Amendment constrains Congress's power to ban material as child pornography. In a First Amendment challenge to a CG-CSAM prosecution under Section 2252A, courts,

---

[32] *United States v. Crisman*, 39 F. Supp. 3d 1189, 1211 (D.N.M. 2014) (citation omitted). The constitutionality of the "morphed images" definition was not before the Court in *Free Speech Coalition*. 535 U.S. at 242. Section 2252A prohibits knowing distribution and production of morphed images with intent to distribute, but not production without intent to distribute, nor (unlike the other categories of CSAM) simple possession. 18 U.S.C. § 2252A(a)(7); U.S. Sentencing Commission, *The History of the Child Pornography Guidelines* (2009): 51 (hereafter *Guidelines History*). At least one morphed-image case has involved a receipt charge. *United States v. Bach*, 400 F.3d 622, 626 (8th Cir. 2005).

[33] My terms are a little imprecise. In referring to "photographic CSAM," I mean to cover video, not just still images, though this paper focuses on the latter. Also, "morphed images" can be *made with* "photographic CSAM": some morphed images "morph" an identifiable minor's nonsexual image onto an actual abuse image of a different child, whereas others morph it onto a pornographic image of an adult.

following *Free Speech Coalition* and *Ferber*, will likely focus their analysis on whether there is harm to any real children. That analysis will differ depending on whether the CG-CSAM is abuse-trained, a morphed image, or photorealistic. Courts will likely decide that so long as it is not abuse-trained or a morphed image, photorealistic CG-CSAM is protected by the First Amendment because it does no direct harm, but that the other two categories do harm actual children and thus are not protected.

For *abuse-trained CG-CSAM*, there is a strong argument under *Free Speech Coalition* and *Ferber* that it can constitutionally be proscribed because its production involved harm to the real children whose abuse imagery was in the training data. For one, producing abuse-trained CG-CSAM arguably does create victims: the generation of new images is predicated on ML training sets containing photographic abuse imagery, whose continued circulation harms anew the children depicted. What's more, we live in an era of voracious data-scraping for ML model training purposes. If there is evidence of traffic in *photographic* CSAM for the purpose of illicitly training ML models to produce *virtual* CSAM, it could be argued that the market for training data creates an economic incentive for production of *photographic* CSAM. "Under either rationale," abuse-trained CG-CSAM has "a proximate link to the crime from which it came."[34]

For *morphed-image CG-CSAM*, the odds are also against the defense. The technology used for creating the morphed image is irrelevant; courts focus on the harm to the identifiable child depicted. For that reason, the majority view is that morphed images are categorically unprotected speech. While the "morphed image" clause was not at issue in *Free Speech Coalition*, multiple federal circuits have subsequently ruled that morphed images fall outside the First Amendment. Just like photographic CSAM, morphed images cause reputational, emotional, and privacy harms to the child depicted, even though no child is abused to create them.[35] The outlier, the Eighth Circuit, held that the First Amendment protects morphed images made without abusing an actual child (that is, neither the image of the identifiable child, nor the sexually explicit image onto which it was morphed, was photographic CSAM). The court nevertheless rejected a defendant's as-applied challenge to a CSAM distribution charge, given the compelling interest in protecting the child depicted from the psychological harms caused by the morphed image's distribution.[36]

---

[34] *Free Speech Coalition*, 535 U.S. at 249–50.

[35] *United States v. Mecham*, 950 F.3d 257, 265 (5th Cir. 2020); *Doe v. Boland*, 698 F.3d 877, 881 (6th Cir. 2012); *United States v. Hotaling*, 634 F.3d 725, 730 (2d Cir. 2011) (distinguishing computer-generated images "where no actual person's image and reputation were implicated"); cf. *Shoemaker v. Taylor*, 730 F.3d 778, 786–88 (9th Cir. 2013) (the Supreme Court has not clearly established free speech protections for morphed images, which are harmful to children just as photographic CSAM is).

[36] *United States v. Anderson*, 759 F.3d 891, 893–96 (8th Cir. 2014) (rejecting the argument that morphed images' harms are "indirect harms" like in *Free Speech Coalition*). Had the defendant not sent the images to

By contrast, *photorealistic CG-CSAM* is likely to be deemed constitutional so long as it involves no direct harm. This category is aptly described by Section 2256(8)(B) as a "computer-generated image that … is indistinguishable from[] that of a minor engaging in sexually explicit conduct."[37] As noted, thanks to overlap in categories, some photorealistic CG-CSAM will either have been trained on photographic CSAM and/or depict an identifiable child. Those images are likely to be deemed illegal for the reasons just stated. However, some photorealistic CG-CSAM will have been produced without photographic abuse imagery in the training data and will not depict an identifiable child. That type of photorealistic CG-CSAM does not directly harm any actual children. What's more, the Court has already rejected many of the arguments for banning photorealistic CG-CSAM: It may cause indirect harms (e.g., inspiring acts of sexual abuse, usage for grooming real children), it may be traded in the same market as photographic CSAM,[38] and defendants may try to evade prosecution by claiming photographic imagery is virtual. Plus, *Williams* confirmed that virtual CSAM is constitutionally protected and the market for it is legal. Under current Supreme Court precedent, then, photorealistic CG-CSAM that is *wholly virtual*—that is, it wasn't trained on photographic abuse imagery and doesn't depict an identifiable child—is protected speech.

In short, whether a particular piece of CG-CSAM may or may not be banned as child pornography is a fact-dependent question turning on whether the training data and/or the resulting image involved direct harm to an actual child.

Constitutionality of CG-CSAM Under Section 1466A

For Section 1466A, the constitutional analysis of CG-CSAM is far more straightforward. That is because the *Free Speech Coalition* decision is irrelevant to Section 1466A. It is a child obscenity statute, not a child pornography statute, so *Miller*, not *Ferber*, is the proper constitutional test.[39] Whether or not the image is photorealistic, depicts an identifiable child, or was trained on photographic abuse imagery, CG-CSAM is not protected speech if it is legally obscene. True, sexually explicit depictions of children will typically meet the *Miller* test, meaning so will most CG-CSAM. However, there might be a small amount of CG-CSAM that Section 1466A cannot constitutionally prohibit.

---

the child depicted in them, he might have escaped prosecution, as Section 2252A does not criminalize simple possession or private production of morphed images. See *Guidelines History* at 51.

[37] 18 U.S.C. § 2256(8)(B).

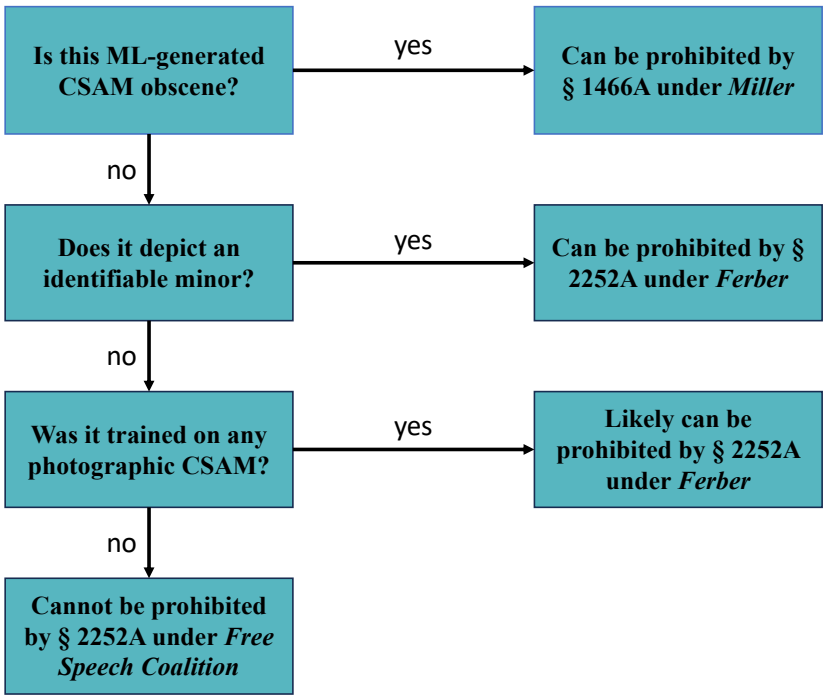
[38] For example, on a Japanese picture-sharing site that contained CG-CSAM, users were also linking to or offering to provide photographic abuse imagery. Angus Crawford & Tony Smith, "Illegal Trade in AI Child Sex Abuse Images Exposed," BBC News, June 28, 2023.

[39] *Mees*, 2009 U.S. Dist. LEXIS 48801, at *11 (*Free Speech Coalition* "did not extend [First Amendment] protection to obscene virtual images").

Two appeals court decisions suggest that some CG-CSAM would not be "patently offensive" under *Miller*. In *United States v. Dean*, the Eleventh Circuit conceded that material depicting "graphic" sexual conduct by minors would usually meet all three *Miller* prongs. Yet the court elucidated a "very narrow" category of First Amendment–protected materials that are nominally prohibited by 1466A: "those that used adult actors *or computer models* to depict older teenagers engaged in non-offensive sexual acts"—specifically, "consenting," "older looking" teenagers.[40] And in fact, the Fifth Circuit recently reversed a 1466A conviction in *United States v. Arthur* because it found not patently offensive a simple drawing of an adolescent girl masturbating that involved no element of force or violence.[41] *Dean* assumed it would be rare for depictions of minors not to be obscene under *Miller*, but *Arthur* shows it does occur. Thus, some wholly virtual CG-CSAM might be not obscene and hence be protected speech.

Summary

Non-obscene CG-CSAM (photorealistic or not) that is wholly virtual—meaning it's not a morphed image and no photographic CSAM was used in the training data—is protected speech under current First Amendment jurisprudence. Prosecutors enforcing child exploitation and obscenity laws, as well as policymakers considering responses to CG-CSAM, must account for the constitutionally significant distinctions among categories of ML-generated material.



[40] *Dean*, 635 F.3d at 1206–08 (emphasis added).

[41] 51 F.4th 560, 570–71 & n.6 (5th Cir. 2022) (noting "the fact-specific nature of the *Miller* test").

IMPLICATIONS FOR REPORTING, PROSECUTION, AND POLICY

It is important to distinguish a constitutional analysis of CG-CSAM from personal opinions about it. The First Amendment protects much speech that many people find repugnant. That is not a weakness or an accident; it is one of the First Amendment's most vital functions. But it means that crafting a response to CG-CSAM requires more nuance than simply banning it all. The CSAM problem is complex enough as it is: The current legal regime is already deluged by the volume of CSAM online. The advent of photorealistic CG-CSAM threatens to further overwhelm the system, confronting law enforcement authorities with complex legal questions and the need to make hard choices among competing priorities.

CG-CSAM Will Further Burden Already-Overwhelmed Reporting Systems

The rise of photorealistic CG-CSAM is predicted to spur "hundreds of thousands of reports to technology platforms, [nongovernmental organizations] handling CSAM cases, and law enforcement, thus overloading the ability of companies and organizations to effectively handle reporting and investigations."[42] This scenario is the foreseeable result of current CSAM law.

Federal law provides that when online service providers discover an "apparent" violation on their services of federal CSAM laws, they must report it to the National Center for Missing and Exploited Children (NCMEC).[43] The reporting requirement covers violations of Section 2252A but not 1466A. That is, platforms don't need to report obscene images such as cartoons that don't document real-life abuse of actual children, but an image must be reported if it "apparently" depicts abuse of a real child.

Now that computer-generated imagery has become less cartoonish and more realistic, CG-CSAM is likely to be reported as an "apparent" 2252A violation. Even if it isn't photorealistic (i.e., less "apparently" illegal), platforms have practical, policy, and legal incentives to take the conservative approach and remove and report CG material to NCMEC rather than attempt to discern whether an image is real or wholly artificial.

On the practical level, there is little reason for a platform to devote resources to making that determination. Platforms sent over 31 million reports to NCMEC in 2022.[44] For platforms to build a reporting pipeline that can operate at scale entails significant investment in personnel and technical systems. Platforms' existing systems were not built for large-scale detection of novel synthetic media. Discerning photographic abuse imagery from photorealistic computer-generated imagery will require new tools. It will be technically difficult and time consuming, and it requires specialized expertise

[42] SIO/Thorn Paper, supra note 1, at 7–8.

[43] 18 U.S.C. § 2258A.

[44] National Center for Missing and Exploited Children, "CyberTipline 2022 Report," https://perma.cc/N8BP-U8H8.

beyond the level of training for platforms' front-line content moderation personnel. Even where those personnel can tell an image is computer generated, they probably lack the information to know if it is a morphed image or was trained on photographic abuse imagery. Platforms are thus likelier to report and remove everything CSAM-esque (and update their terms of service to ban CG-CSAM) than invest resources in discerning CG-CSAM from photographic CSAM.

As a policy matter, platforms may have little appetite for drawing such a distinction anyway. Failing to accurately identify and remove actual photographic CSAM—or merely appearing cavalier about efforts to do so—can be a public relations and government relations nightmare. Meta suffered a negative press cycle in 2022 following press coverage of a corporate training document that taught content moderators to "err on the side of an adult" and not report an image when the moderator was not sure if the person depicted was under age.[45] Recent research finding widespread "self-generated" CSAM on Instagram and Twitter prompted significant media coverage and congressional interest.[46] Even a single mistaken failure to remove photographic CSAM can make headlines.[47] For this reason as well, the incentive is to report CG-CSAM, photorealistic or not.

Finally, legal considerations also militate strongly in favor of reporting CG-CSAM (and banning it under terms of service). Failing to report photographic CSAM risks criminal liability and significant fines, whereas platforms are immune from liability (with limited exceptions) if they report or remove material that does *not* qualify as CSAM.[48] There are major legal risks in not reporting something that's actually CSAM and little to no legal risk in reporting something that's not actually CSAM.

Unsurprisingly given these factors, platforms already report material they aren't required to. NCMEC's reporting API lets platforms annotate a file as "depicting anime, drawing, cartoon, virtual or hentai," even though those are expressly excluded from the definition of CSAM.[49] Platforms also sometimes

[45] Michael H. Keller, "Adults or Sexually Abused Minors? Getting It Right Vexes Facebook," *New York Times*, Mar. 31, 2022.

[46] David Thiel, Renée DiResta, and Alex Stamos, "Cross-Platform Dynamics of Self-Generated CSAM," Stanford Internet Observatory, June 2023, https://perma.cc/3939-SPYV; "Meta Asked by Senators to Detail Efforts to Prevent Sharing of Child Sexual-Abuse Material," *Wall Street Journal*, June 30, 2023.

[47] Gabrielle Fonrouge, "Twitter Refused to Remove Child Porn Because It Didn't 'Violate Policies': Lawsuit," *New York Post,* Jan. 21, 2021.

[48] 18 U.S.C. §§ 2258A(e), 2258B; 47 U.S.C. § 230.

[49] CyberTipline Reporting API Technical Documentation, National Center for Missing and Exploited Children (n.d.), https://perma.cc/EQE2-8U4D; 18 U.S.C. § 2256(11).

report nonsexual child nudity images such as photos a parent emailed to a pediatrician, which can severely disrupt innocent people's lives.[50]

In sum, federal law incentivizes platforms to externalize onto NCMEC and law enforcement the task of determining whether a reported image involves harm to an actual child. Once reports are submitted, it is up to NCMEC to divide "actionable" tips from "informational" ones, a category that includes tips with "no apparent child sexual exploitation nexus."[51] Dealing with such reports wastes analysts' time and organizational money.[52] As CG-CSAM becomes more commonplace, it will further burden platforms' reporting pipelines on the front end and use up even more resources on the receiving end. By worsening the already poor signal-to-noise ratio, CG-CSAM risks impairing investigators' ability to efficiently identify and intervene when a real child is being harmed.

How Will Prosecutors Deal With the Advent of Photorealistic CG-CSAM?

Photorealistic CG-CSAM will present prosecutors with tough dilemmas in choosing which cases to prosecute and how, given limited resources, competing priorities, larger policy considerations, and constitutional constraints.

Evidence Will Play a Paramount Role in Prosecuting CG-CSAM

Given the serious challenges in prosecuting CG-CSAM, it is foreseeable that prosecutors will pick their battles carefully. The potential for a successful First Amendment challenge under *Free Speech Coalition* or *Miller* disfavors prosecution of photorealistic CG-CSAM that is shown to be wholly virtual or might not be deemed obscene. By contrast, prosecutors can be more confident of prevailing where they can show that an image is photographic and not CG, that CG-CSAM depicts an identifiable child,[53] or that a

[50] Kashmir Hill, "How Your Child's Online Mistake Can Ruin Your Digital Life," *New York Times*, Nov. 27, 2023; Kashmir Hill, "A Dad Took Photos of His Naked Toddler for the Doctor. Google Flagged Him as a Criminal," *New York Times*, Aug. 21, 2022.

[51] U.S. Department of Justice Office of Justice Programs, "CY 2022 Report to the Committees on Appropriations," https://perma.cc/N8BP-U8H8.

[52] Internet Watch Foundation, "'We've Got to Get a Grip on the Epidemic on Our Open Internet' - UK Charity Deals With Record Number of Reports of Online Child Sexual Abuse Material," Jan. 6, 2020, https://perma.cc/CQK3-KEU7.

[53] For example, the teenage girls already being depicted in pornographic deepfakes are identifiable minors under the morphed-image law. Haleluya Hadero, "Teen Girls Are Being Victimized by Deepfake Nudes. One Family Is Pushing for More Protections," *Seattle Times*, Dec. 2, 2023.

CG image was generated using photographic abuse imagery. The outcome of those cases will turn on evidentiary issues, not constitutional ones.

Once photorealistic CG-CSAM becomes common, it may not be clear at the outset what category an image falls into: photographic, wholly virtual, abuse-trained, or morphed-image. The parties will need to marshal evidence to answer that question. At the beginning of the twenty-first century, imaging technology was still rudimentary enough that proving an image was produced using real children (as *Free Speech Coalition* requires) didn't pose an obstacle. With the advent of diffusion models, that determination may become more difficult.

In addition, the prosecution must prove every element of an offense beyond a reasonable doubt. While it is an affirmative defense to most charges under Section 2252A (but not 1466A) that no real child was involved,[54] the *Free Speech Coalition* Court was skeptical that the law could constitutionally shift the burden of proof onto the defendant "to prove the speech is lawful."[55] Since *Free Speech Coalition*, courts have repeatedly held that the government must prove beyond a reasonable doubt that the child depicted is real.[56] A defendant's failure to raise the affirmative defense does not relieve the government of its burden of proof.[57] That said, courts have refused to impose a heightened burden of proof to establish this element, such as the use of expert testimony.[58] Indeed, the government need not introduce any extrinsic evidence beyond the images.[59]

The advent of truly photorealistic CG-CSAM cannot relieve the government of its due process obligations. Even if the defendant never invokes the affirmative defense, the prosecution must prove to the jury's satisfaction that an image was produced using real children. While the government isn't *required* to introduce expert testimony, it has often done so anyway, and as a practical matter that may be what it takes to persuade a jury in the coming era of photorealistic CG-CSAM. Expert witnesses in digital forensics and related fields will be in high demand by both sides to testify about whether an accused image is photographic or virtual. If the image's underlying training data could be ascertained, a fact witness might also testify to whether the data included photographic abuse imagery. (If the

---

[54] 18 U.S.C. § 2252A(c)(2). The defense is not available for morphed-image charges. Id.

[55] 535 U.S. at 255–56.

[56] *United States v. Salcido*, 506 F.3d 729, 733 (9th Cir. 2007) (per curiam) (collecting cases); *United States v. Rodriguez-Pacheco*, 475 F.3d 434, 439 (1st Cir. 2007); *Irving v. United States*, 452 F.3d 110, 122 (2d Cir. 2006).

[57] *United States v. Hilton*, 386 F.3d 13, 18 (1st Cir. 2004) (per curiam).

[58] *United States v. McNealy*, 625 F.3d 858, 865 & nn. 24–30 (5th Cir. 2010) (collecting cases); *United States v. Lacey*, 569 F.3d 319, 325 (7th Cir. 2009); *Rodriguez-Pacheco*, 475 F.3d at 441.

[59] *United States v. Bynum*, 604 F.3d 161, 166 (4th Cir. 2010).

generated image depicts a real victim's face, that will strongly indicate that it was trained on abuse imagery of that victim, or that the victim's face had been morphed into the resulting image.)

None of this means that the advent of technology for creating truly photorealistic CG-CSAM will stymie the government's ability to obtain convictions in cases involving photographic CSAM. This "prosecutorial necessity" argument was the government's rationale to justify banning wholly virtual CSAM that did no direct harm to any real child, and the *Free Speech Coalition* Court rejected it, saying it "turns the First Amendment upside down."[60] That was in an earlier era of computer imaging technology. But there are strong reasons to believe that a prosecutorial crisis will still not materialize even in the new age of truly photorealistic CG-CSAM. That means "prosecutorial necessity" still doesn't justify stripping wholly virtual CSAM of its First Amendment protection (even though that is arguably what the PROTECT Act tried to do with its untested "indistinguishable" definition of CSAM).

Key to the "prosecutorial necessity" argument is the notion that traffickers of photographic CSAM "may escape conviction by claiming that the images are computer generated, thereby raising a reasonable doubt as to their guilt."[61] In reality, though, this defense has been a Hail Mary pass that rarely lands—as Justice Clarence Thomas noted in his *Free Speech Coalition* concurrence and as Justice David Souter meticulously documented in his lengthy dissent in *Williams*.[62] If a defendant asserts a "computer-generated images" defense, creating reasonable doubt in the jury's minds will take more than just pointing to the general existence of the technology for making photorealistic virtual images without using actual children. That is, it's not enough to argue that diffusion models exist and so it's possible the images in question were created that way. Rather, the defendant must provide some evidence that the specific images at issue in *his* case are computer generated.[63] A defendant is "free to … present[] evidence of his own suggesting that the pictures" don't depict real children, but mere conjecture won't cut it.[64]

---

[60] 535 U.S. at 254–55.

[61] *Free Speech Coalition*, 535 U.S. at 259 (concurrence of Thomas, J.).

[62] Id.; *Williams*, 553 U.S. at 323–26 & nn. 3–4 (Souter, J., dissenting).

[63] The government's burden is equally concrete: It cannot rely on "the law of averages" by arguing that the accused images must be of real children because most abuse imagery in circulation is; it must prove that those particular images depict real children. *United States v. Wilder*, 526 F.3d 1, 14 (1st Cir. 2008) (Stahl, J., concurring).

[64] *United States v. Nolan*, 818 F.2d 1015, 1019–20 (1st Cir. 1987) (rejecting "uncorroborated speculation that some undefined technology exists to produce these pictures without use of real children"); *United States v. Fuller*, 77 Fed. Appx. 371, 380 (6th Cir. 2003) (unpub.) (jury could "draw its own conclusions about whether [images] depicted real children" where the defendant offered "no contrary evidence … to suggest … that any of the visual depictions were computer generated") (citations omitted).

Courts have long held that the government's burden does not require it "to negate unsupported speculation that images may have been computer-generated or other than what they appeared to be."[65] If a defendant chooses to raise this defense and goes beyond baseless conjecture to present actual evidence, such as expert testimony, to support it, then "the government [runs] the risk of not persuading the trier of fact" if it fails to answer that evidence with its own, since the government retains the burden of proof at all times.[66] But carrying that burden does not require the government "to rule out every conceivable way the pictures could have been made other than by ordinary photography."[67]

This is why the "computer-generated images" defense has routinely failed. It wasn't just that until very recently, the state of the art was too rudimentary or too expensive for that defense to hold much water. It's that even if the technology existed, defendants didn't do enough to connect it to the images in *their* case. The rise of ML diffusion models won't suddenly give teeth to a historically empty defense. Yes, the defense might get raised even more frequently. But without something more to back it up, mere speculation still won't suffice to establish reasonable doubt. That was true when *Free Speech Coalition* came out and is still true twenty years later.[68]

Even if this defense does start to succeed sometimes, it is questionable whether that can justify regulating wholly virtual CSAM or even removing it entirely from the First Amendment's aegis. In the 2000s, multiple Supreme Court justices indicated that such measures might become acceptable if technology really did start thwarting prosecutions of those who possess and disseminate "real" CSAM.[69] Justice Thomas, for example, speculated that future technological advances might one day stymie CSAM prosecutions "because the Government cannot prove that certain pornographic images are of real

[65] *United States v. Deaton*, 328 F.3d 454, 455 (8th Cir. 2003) (citing *United States v. Vig*, 167 F.3d 443, 449–50 (8th Cir. 1999)).

[66] *Rodriguez-Pacheco*, 475 F.3d at 444.

[67] *Nolan*, 818 F.2d at 1020 (citation omitted); *Rodriguez-Pacheco*, 475 F.3d at 439 (reaffirming post-*Free Speech Coalition* the continuing vitality of *Nolan*'s rule "that the mere possibility, unsupported by evidence, that the images could have been produced by use of technology and not using real children [is] not sufficient" for a reviewing court to reject the lower court's conclusion that the images were real).

[68] For example, *State v. MacHardy*, 521 P.3d 613, 624–25 (Ariz. Ct. App. 2022) (The state introduced sufficient evidence that images depicted actual children, whereas the defendant presented no evidence supporting his argument about "advances in technology" enabling digital alteration of images; "the state is not required to rule out every possibility of innocence, no matter how speculative, to be entitled to a guilty verdict[.]") (citation omitted), *review denied*, 2023 Ariz. LEXIS 129 (Ariz. 2023).

[69] *Free Speech Coalition*, 535 U.S. at 259 (concurrence of Thomas, J.); id. at 263–64 (concurrence in part and dissent in part of O'Connor, J.); *Williams*, 553 U.S. at 323 (dissent of Souter, J.).

children." In that event, he said, "barring or otherwise regulating" virtual CSAM might be necessary for the government to effectively enforce the law in cases involving real-life child abuse.[70]

But what does it really mean to say that regulating or even banning currently protected depictions of virtual abuse will become permissible if prosecutors cannot prove that certain images are actually unprotected depictions of real abuse? This proposal sounds a lot like reducing the scope of one constitutional right—freedom of speech—to undercut the consequences of other rights to due process and against double jeopardy.

In a system that presumes innocence, it is the government's burden, always, to prove that alleged CSAM is photographic and that the defendant is guilty. If a defendant introduces evidence in support of the "computer-generated images" defense that, without adequate rebuttal by the government, is sufficient to raise a reasonable doubt in the minds of the jury that the accused images involve real children, then the government has failed to carry its burden of proof, the jury must acquit, and the government cannot appeal the acquittal.

That's the trouble with arguing that if the "computer-generated images" defense ever starts to prevail (which it hasn't), "prosecutorial necessity" justifies a new law restricting virtual CSAM. Claiming the law is necessary because prosecutors failed to convict in "real" CSAM cases means disregarding juries' not-guilty verdicts and assuming the very things the government could not prove: that acquitted defendants were guilty of exploiting real children. This approach would treat acquittals by criminal juries as a systemic failure to be rectified through other means, namely the enactment of content-based restrictions on protected speech. It does an end-run around the Fifth Amendment to propose that if the government can't prove beyond a reasonable doubt that an image is unlawful speech rather than lawful speech, then the answer is to make the lawful speech unlawful too.

For decades, as imaging technology has evolved, CSAM defendants' baseless speculation about it has failed to yield acquittals. There is no reason so far to believe this time will be different. As Justice Souter wrote fifteen years ago, "experience tells us to have faith in the capacity of the jury system" before manufacturing an "excuse for cutting back on the First Amendment."[71] Photorealistic CG-CSAM's mere invention does not create the "prosecutorial necessity" to revisit *Free Speech Coalition*'s rule that virtual CSAM is protected speech.[72]

---

[70] *Free Speech Coalition*, 535 U.S. at 259 (Thomas, J., concurring).

[71] *Williams*, 553 U.S. at 326–27 (Souter, J., dissenting).

[72] Of course, the Court's composition has changed radically since 2002. Only Justice Thomas remains, sitting as part of a hard-right majority that has proved willing to overturn settled precedent about Americans' constitutional rights. See *Dobbs v. Jackson Women's Health Org.*, No. 19-1392, 597 U.S. ___ (2022) (rescinding the 49-year-old right to abortion). The current Court might thus be open to overruling *Free Speech Coalition* and narrowing the First Amendment's scope, but the thought is not exactly comforting.

The Drawbacks of Relying on Obscenity Law Instead of CSAM Law

It is unconstitutional to prosecute wholly virtual material as CSAM, but there might be major evidentiary hurdles in determining whether an accused CG-CSAM image is wholly virtual. By contrast, proving the involvement of a real child is irrelevant to the child obscenity statute. That might make it seem like an attractive alternative,[73] but it has its own problems.

Relying on Section 1466A would mean prosecuting only CG-CSAM that is obscene. That could be only a slightly smaller universe than what Section 2252A covers.[74] But the remainder poses a quandary for prosecutors: Proceeding under 2252A risks getting the untested "virtually indistinguishable" definition of CSAM struck down as unconstitutional; going with 1466A risks losing on the merits of the *Miller* test. The government might thus prioritize prosecuting CG-CSAM of younger children as obscene over imagery depicting older teenagers.

Shifting to greater use of Section 1466A would also implicate broader societal issues. Because the *Miller* test lets local juries decide "community standards," increased reliance on obscenity law risks enshrining regressive social norms about sex, sexuality, and sexual orientation. Courts and juries considered same-sex content obscene well into the 1990s.[75] We're now once more in an age of widespread anti-LGBTQI+ prejudice that equates queer people with "child groomers." Therefore, a jury's decision about CG-CSAM's obscenity might hinge on whether or not the image depicted heterosexual conduct and cisgender people. Using 1466A against CG-CSAM that presents a close question under *Miller* could exacerbate the anti-LGBTQI+ backlash by letting jurors' biases dictate whether someone goes to prison. In those circumstances, a win for the government is not a win for society.

Triaging CG-CSAM Prosecutions

If CG-CSAM becomes commonplace, prosecutors will have to make tough decisions. As always, not every potential case will be prosecuted given limited resources. Resource constraints may be particularly acute given the sophisticated digital forensics and experts that will be needed in many cases. The cost

---

[73] Already, federal obscenity prosecutions usually involve CSAM or alleged harm to minors. Jennifer Kinsley, "The Myth of Obsolete Obscenity," *Cardozo Arts & Entertainment Law Journal* 33 (2015): 607, 637.

[74] See *Dean*, 635 F.3d at 1207–08 (only imagery of older-looking, consenting teenagers engaging in "non-offensive" sex acts might fail *Miller*).

[75] Kendra Albert, "Imagine a Community: Obscenity's History and Moderating Speech Online," *Yale Journal of Law & Technology* Special Issue 25 (2023): 59, 72 (citing *Tipp-It, Inc. v. Conboy*, 596 N.W.2d 304 (Neb. 1999), as finding "material that would be unlikely to be scandalous if [it] involved male and female participants to be legally obscene when involving two men").

and complexity of this technical challenge might weigh into a prosecutor's decision to pursue charges—or, for the defendant, to plead out.

Another factor is whether a real child was harmed, and the degree of the harm.[76] Resource constraints can cut at cross-purposes to this factor: Harm to a real child matters in 2252A cases but not 1466A cases, so an obscenity case might be less resource intensive to prosecute than a 2252A case. That said, CG-CSAM trained on imagery of actual child abuse should take prosecutorial priority over wholly virtual obscene material, even if the prosecution is costlier.

Another potential factor is whether the defendant's activities contributed to the proliferation of abuse-trained CG-CSAM. That is, prosecutors might choose to focus on alleged producers and disseminators of such material (especially at large scale) over those who simply possess it. Here there is more alignment between real-world harm and resource constraints: It should be more feasible to confirm that an accused producer trained an ML model on photographic abuse imagery, compared with possession cases involving images downloaded from the internet.

Taken together, these factors might lead, for example, to different prioritization of the generation of illegal CG-CSAM for large-scale distribution or sale versus generation for purely private use (which has constitutional problems of its own, as discussed later). Distribution of morphed-image CG-CSAM depicting an identifiable child might be prioritized differently than possession of CG-CSAM whose underlying training data cannot be determined. And so on. Such triage will be an ugly necessity in the coming era of photorealistic CG-CSAM.

There is another way to bring CG-CSAM offenders to justice: look for their photographic CSAM. Even if ML diffusion models lead to a flood of new CG-CSAM, known and new CSAM of real children will continue to circulate. CSAM offenders often possess extensive troves of illegal material.[77] If demand for CG-CSAM is intermingled with the market for photographic CSAM, then those who possess the former may possess the latter too. If an investigation reveals both virtual CSAM and photographic CSAM, prosecutors can, and do, bring charges for the latter only.[78] Not only does this approach to photorealistic

[76] The Department of Justice already draws this distinction; it considers morphed-image production "not as serious a crime as the production of genuine [CSAM]." *Guidelines History* at 50–51.

[77] "Many offenders possess child pornography collections numbering in the hundreds of thousands or even millions of images and videos." U.S. Sentencing Commission, "2012 Report to the Congress: Federal Child Pornography Offenses," at 45 (citation omitted).

[78] For example, *United States v. Knott*, No. 21-cr-328, 2022 U.S. Dist. LEXIS 198553, at *2, *6 (M.D. Ala. Nov. 1, 2022) (defendant was charged with one count of possessing photographic CSAM; although he also possessed over 100,000 computer-generated images, the government did not allege they constituted CSAM); see also *Lacey*, 569 F.3d at 325 & n. 6 (affirming sentencing enhancement for possession of more than 600 CSAM images, where defendant unsuccessfully challenged the court's finding that still images were of actual children but did not dispute that videos were of actual children). This can also be a viable option in

CG-CSAM avoid difficult constitutional and evidentiary issues, but it also focuses prosecutorial resources on instances of definite harm to real children.

If CG-CSAM triggered a police search that turned up photographic CSAM, the defense might challenge the constitutionality of the search. There are First and Fourth Amendment concerns with letting *legal* speech supply probable cause to search for *illegal* material. Legal images, without more, might not be enough for probable cause.[79] But probable cause can be based on innocent behavior; plus, the more photorealistic the image, the less innocent it looks and the more like evidence of a crime.[80] In any event, judges might be loath to refuse to issue a warrant in such circumstances, or to invalidate a search on appeal. Whether this dynamic is a net positive for society is open for debate.[81]

Potential Policy Responses

Addressing the rise of CG-CSAM will require policy interventions from a range of stakeholders. While measures for the creators of diffusion models to implement are beyond the scope of this paper,[82] there is a role for government to play. While the First Amendment limits the options available, Congress's hands are not wholly tied from enacting legislation to prevent the proliferation of abuse-trained CG-CSAM. This section suggests a few options.

Prohibiting Trafficking in Abuse-Trained ML Tools

One place to start is by targeting the tools for producing abuse-trained CG-CSAM. Current CSAM law does not outlaw the instrumentalities used for producing CSAM. This makes sense given that those tools have historically been general purpose: Mere ownership of a digital camera or Photoshop software is not

---

states whose CSAM law does not cover morphed images. See Tiffany Hsu, "What Can You Do When A.I. Lies About You?" *New York Times*, Aug. 3, 2023 (creator of pornographic deepfakes of teen girls was convicted under state law for photographic CSAM he had).

[79] Cf. *United States v. Hansel*, 524 F.3d 841, 846 (8th Cir. 2008).

[80] See *Illinois v. Gates*, 462 U.S. 213, 243 n.13 (1983) (establishing "totality of the circumstances" test for probable cause).

[81] See *United States v. Bosyk*, 933 F.3d 319, 370 (4th Cir. 2019) (Wynn, J., dissenting) ("though the vileness of the crime in child pornography cases often presents a strong incentive to take liberties with the protections the Constitution affords to criminal defendants, we must guard our roles as judges to resist even that temptation") (citing *United States v. Coreas*, 419 F.3d 151, 151 (2d Cir. 2005)); *United States v. Gourde*, 440 F.3d 1065, 1077–78 (9th Cir. 2006) (Kleinfeld, J., dissenting) ("Sex with children is so disgusting to most of us that we may be too liberal in allowing searches when the government investigates child pornography cases.").

[82] For a range of suggestions, see the SIO/Thorn Paper, supra note 1, at 9–16.

illegal; what is criminalized is the particular use of that tool. The distinction is one that Congress has recognized in other contexts: To avoid criminalizing general-purpose technology, Congress has enacted statutes that prohibit the manufacture, possession, and trafficking of tools whose primary or sole functionality is illegal, such as eavesdropping on communications or the circumvention of digital rights management on copyrighted works.[83]

The phenomenon of ML model training reveals an apparent gap in existing CSAM law: It does not encompass the possession or dissemination of ML models that were trained on photographic abuse imagery. LoRA files, for example, are not image files; they contain algorithms, not the images on which they were trained. Congress could amend the federal child sex abuse and exploitation statutes to add a narrowly drawn provision addressing the knowing and intentional misuse of ML tools for CG-CSAM purposes. While such a statute would likely draw a constitutional challenge, it might pass muster for the same reasons discussed above regarding abuse-trained CG-CSAM. That is, abuse-trained ML tools, while they do not themselves contain abuse imagery, nevertheless have "a proximate link to the crime[s]" of CSAM production, possession, and dissemination.[84]

It would be imperative to include carefully tailored *mens rea* requirements and carve-outs to avoid capturing innocent conduct, such as possession, sharing, or use of an ML model without knowing or having reason to know that it had been trained on photographic abuse imagery. That is not a hypothetical situation: Recent research revealed that the "LAION-5B" ML data set, popular for training AI generation models, contained links to at least 1,000 instances of known CSAM.[85] In response, the organization responsible for the data set took it down, but the horse has fled the barn: Copies of the data set had already been made and presumably are in wide private possession.[86] This is why *mens rea* requirements and fair notice of what the law prohibits are crucial in criminal law. Surely everyone who downloaded or shared a copy of LAION-5B (which, as said, contains links, not images) is not culpable under current CSAM law. Nevertheless, this incident demonstrates that ML models themselves, not just their outputs, will require policymakers' attention.

---

[83] 18 U.S.C. § 2512 (eavesdropping); 17 U.S.C. § 1201(a)(2), (b)(1) (circumvention).

[84] *Free Speech Coalition*, 535 U.S. at 250.

[85] David Thiel, "Identifying and Eliminating CSAM in Generative ML Training Data and Models," Stanford Internet Observatory, Dec. 21, 2023, https://perma.cc/3939-SPYV (hereafter LAION-5B Report). "LAION datasets do not include the actual images; instead, they include a link to the original image on the site from which it was scraped." Id. at 5.

[86] Samantha Cole, "Largest Dataset Powering AI Images Removed After Discovery of Child Sexual Abuse Material," 404 Media, Dec. 20, 2023, https://perma.cc/PEU8-P5ZB (quoting Thiel as saying, "If you have downloaded that full dataset for whatever purpose, for training a model for research purposes, then yes, you absolutely have CSAM, unless you took some extraordinary measures to stop it").

Leaving aside the legality of LAION-5B itself, how should the law treat all the images that resulted from LAION-5B, for which the presence of CSAM likely influences the model's outputs?[87] In my analysis, it is likely constitutional to criminalize ML-generated images *of child sex abuse* where the ML model was trained on actual abuse imagery. But what about abuse-trained, ML-generated images that do *not* depict child sex abuse—for example, images made with LAION-5B of adult pornography, or that are not sexual at all? How far does the First Amendment allow *Ferber*'s "harm to real children" rationale to extend when it comes to ML-generated imagery that, to look at it, has no connection to children or sex? This is a new conundrum, born of ML models' stark differences from previous generations of technology for creating or manipulating images. Given the significant constitutional issues with prosecuting non-abuse-depicting ML-generated images that may have photographic CSAM somewhere in their metaphorical DNA, the legal questions about LAION-5B and the images made with it are less pressing than the practical dilemma of what to do about them, which has no easy answers.[88]

Including CG-CSAM in the National Strategy for Synthetic Imagery

Another priority area for legislation is image authentication. In passing the PROTECT Act, Congress was concerned about the potential negative impact on prosecutions of "limiting the child-pornography prohibition to material that could be *proved* to feature actual children, as [the] decision in *Free Speech Coalition* required."[89] If Congress believes the problem is the difficulty of telling photographic and virtual abuse imagery apart, then its top priority should be to address that technical issue, not the passage or enforcement of constitutionally dubious laws.

There is an urgent need, exacerbated by the breakneck pace of advancements in machine learning, for Congress to invest in solving this technical challenge. Government investment could also aid the development of technology to counteract modifications to photographic abuse imagery that might otherwise impede investigation, such as manipulations that obscure an image's setting or make a photographic image look like the output of a diffusion model. To that end, in its recent executive order on the "safe, secure, and trustworthy development and use" of AI, the Biden administration included multiple provisions pertaining to the prevention of CG-CSAM as part of its larger strategy for reducing the risks posed by synthetic content.[90]

The problems are not confined to CSAM: As the AI executive order recognizes, synthetic or manipulated imagery presents a larger societal challenge of which CG-CSAM is merely one part. In 2020, Congress enacted a "deepfakes" law directing the National Science Foundation (NSF) and the

[87] LAION-5B Report, supra note 85, at 11.

[88] See LAION-5B Report, supra note 85, at 11–12 (identifying multiple aspects of the problem, "of increasing difficulty").

[89] *Williams*, 553 U.S. at 290 (citing § 501(9), (10), 117 Stat. 677).

[90] Executive Order No. 14110, 88 Fed. Reg. 75,191 (Nov. 1, 2023).

National Institute of Standards and Technology (NIST) to support research into, and develop standards for detecting, synthetic or manipulated media.[91] Other bills relating to synthetic media detection and digital media provenance have also been proposed.[92] Relatedly, the Defense Advanced Research Projects Agency (DARPA) has had two programs concerning deepfake detection, MediFor (concluded in 2021) and SemaFor (ongoing).[93]

Congress, like the White House, should make CG-CSAM part of its overall strategy for synthetic imagery issues. It could examine the NSF, NIST, and DARPA results to date (as well as the upcoming reports required by the AI executive order) and craft a proposal tailored to the CG-CSAM context. Unlike most imagery, possession of CSAM is a crime, with narrow exceptions limiting the liability of certain entities such as NCMEC.[94] Depending on the details, research into CG-CSAM detection could require government approval,[95] which would likely be a contested topic.

The fruits of any (taxpayer-supported) government investment in improved detection and provenance capabilities must not be confined to NCMEC and law enforcement. Internet platforms, ML model creators, academics, and child-safety nongovernmental organizations all could benefit—and, importantly, so would criminal defendants. The accused, who are innocent until proven guilty, have constitutional rights to have their alleged guilt established beyond a reasonable doubt and to confront the evidence against them in a fair trial. In court, "it will be at least as difficult" for defendants to resolve the evidentiary issue of whether an image is (or was trained on) photographic abuse imagery as for the government.[96] Yet public defenders are vastly under-resourced compared to prosecutors.[97] This gap

[91] Identifying Outputs of Generative Adversarial Networks Act (IOGAN Act), 134 Stat. 1150, Pub. L. 116-258 (2020).

[92] For example, Deepfake Task Force Act, S.2559 (117th Cong.).

[93] DARPA, Media Forensics (MediFor) (Archived) (n.d.), https://perma.cc/22HQ-GTZV; DARPA, Semantic Forensics (SemaFor) (n.d.), https://perma.cc/UKH5-4TEK.

[94] See 18 U.S.C. §§ 2258A, 2258B, 2258D.

[95] Drew Harwell, "AI-Generated Child Sex Images Spawn New Nightmare for the Web," *Washington Post*, June 19, 2023 (quoting AI researcher Margaret Mitchell); cf. Cole, supra note 86 (describing the difficulties that legal restrictions pose for researchers who study CSAM).

[96] *Free Speech Coalition*, 535 U.S. at 255–56.

[97] For example, California Legal Analyst's Office, "Assessing the Provision of Criminal Indigent Defense," Sept. 22, 2022, https://perma.cc/ETW2-RVD5; Brian Lee, "Public Defenders Lobby for Funding Parity With Prosecutors," *New York Law Journal*, Mar. 13, 2023.

extends to digital forensics tools and expertise, which can play a crucial role in exoneration.[98] Policymakers should ensure that policy responses to CG-CSAM do not tilt the playing field further against criminal defendants.

Other Potential Policy Responses

Opening governmental purse strings is no easy matter, and yet it may be a relatively agreeable policy intervention. Other options present tougher dilemmas, would be more controversial, and/or raise constitutional concerns.

One top priority that should be uncontroversial is allocating additional funding to NCMEC. As discussed, NCMEC is already overwhelmed by incoming CSAM reports (as are the law enforcement agencies to which it routes them),[99] and CG-CSAM will only increase the load. Additional funding would let NCMEC hire more personnel (including contractors), upgrade its tooling, and expand its use of other resources (such as mental health initiatives).

Alternatively, to take some of the load off NCMEC, Congress could exempt CG-CSAM from platforms' reporting requirements under Section 2258A, or at least add a safe harbor if the platform had a good-faith belief that an unreported image was wholly virtual. This would lessen the incentive for platforms to inundate NCMEC with CG-CSAM reports. But Congress would probably be reluctant to loosen platforms' CSAM-related obligations: The trend of proposed child-safety legislation in recent years has gone firmly in the other direction.[100] What's more, there may be little desire to have platforms' employees try to figure out which images are computer generated, particularly as that determination becomes ever more difficult.

Other measures in response to CG-CSAM would be of questionable constitutionality. For one, while image provenance is generally a worthwhile area for policy investment, there are serious constitutional issues with *requiring* the labeling of wholly virtual images. Labeling, watermarking, or attestation could affirm that an image is wholly virtual and thus, if no photographic CSAM was used in the training, legal under *Free Speech Coalition*. (Whether it is obscene under *Miller* cannot be defined by a label or watermark.) However, such measures would need to be voluntary, as there are serious First Amendment

[98] Kashmir Hill, "Imagine Being on Trial. With Exonerating Evidence Trapped on Your Phone," *New York Times*, Nov. 22, 2019.

[99] For example, Michael H. Keller & Gabriel J.X. Dance, "The Internet Is Overrun With Images of Child Sexual Abuse. What Went Wrong?" *New York Times*, Sept. 29, 2019; Hugh Cook, "Tips on Internet Crimes Against Wyoming Minors Have Risen Dramatically in Recent Years," Wyoming Public Radio, Mar. 9, 2023.

[100] For example, EARN IT Act, S.1207 (118th Cong.); REPORT Act, S.474 (118th Cong.); STOP CSAM Act, S.1199 (118th Cong.).

concerns with mandating labels for protected expression that does not harm any real person.[101] Even if voluntary, there are practical questions, both as to the viability of novel labeling or watermarking technology and about whether anyone would rely on it, given potential legal risks.

What's more, where CG-CSAM creators keep their material to themselves and never share it, they might be protected by the constitutional right to privately possess obscene matter.[102] That right applies only to obscenity, not (photographic) CSAM, given the state's interest in child protection.[103] And it doesn't cover obscene materials that have moved in interstate commerce, like internet downloads.[104] But it means not all private CG-CSAM possession could be banned. Abuse-trained CG-CSAM surely can be, as can obscene CG-CSAM that was downloaded from the internet or generated on third-party servers.[105] But for morphed-image CG-CSAM, the "harm to child victims" rationale carries less force if the image is never shared and the child depicted is unaware of it.[106] And the closest question involves obscene,

---

[101] See Bradley Waldstreicher, "Deeply Fake, Deeply Disturbing, Deeply Constitutional: Why the First Amendment Likely Protects the Creation of Pornographic Deepfakes," *Cardozo Law Review* 42 (2021): 729 (analyzing labeling and watermarking mandates in federal and state proposals to regulate deepfakes); see also *Free Speech Coalition v. Attorney General*, 974 F.3d 408 (3d Cir. 2020) (partially successful First Amendment challenge to 18 U.S.C. §§ 2257 & 2257A, which impose age verification, recordkeeping, and labeling requirements on pornography producers in order to combat CSAM).

[102] *Stanley v. Georgia*, 394 U.S. 557, 559, 568 (1969); *Handley*, 564 F. Supp. 2d at 1000 ("The limited right to possess obscene materials in the privacy of one's own home recognized in *Stanley* depended not on First Amendment grounds, but on the right to privacy in the home found in the Fourth Amendment.") (citing *United States v. 12 200-Foot Reels of Super 8mm. Film*, 413 U.S. 123, 126 (1973)).

[103] *Osborne v. Ohio*, 495 U.S. 103, 108–11 (1990).

[104] *Whorley*, 386 F. Supp. 2d 693, 695 (E.D. Va. 2005) (1466A(a)(1) case involving receipt of materials from the internet). This is why charges for possession of child obscenity (which requires an interstate commerce nexus) typically survive constitutional challenge. 18 U.S.C. § 1466A(d); *United States v. Taylor*, No. ACM 38700, 2016 CCA LEXIS 108, at *22–26 (A.F.C.C.A. Feb. 25, 2016) (unpub.); *Mees*, 2009 U.S. Dist. LEXIS 48801, at *12–13; *Handley*, 564 F. Supp. 2d at 1000–01; cf. *Whorley*, 550 F.3d at 332–33 (rejecting challenge to 18 U.S.C. § 1462); but see *United States v. Gendron*, No. 08-cr-00244, 2010 U.S. Dist. LEXIS 16055, at *7–8 (E.D. Mo. Feb. 23, 2010) (inviting briefing on constitutional issue; government then dropped the charge).

[105] Third-party services with insufficient guardrails against CG-CSAM are becoming more cognizant of their potential liability and taking measures to clamp down on this misuse of their services. Emanuel Maiberg, "Civitai and OctoML Introduce Radical New Measures to Stop Abuse After 404 Media Investigation," 404 Media, Dec. 9, 2023, https://perma.cc/W5W5-KZRY.

[106] This may be why Section 2252A does not already criminalize simple possession of morphed images or production thereof without intent to distribute them. See *Guidelines History* at 51.

wholly virtual CG-CSAM generated at home using *equipment* that had moved in commerce.[107] Section 1466A's current Commerce Clause language falls short of "the outer limits of [Congress's] authority."[108] But there must be some limit; otherwise, stretched far enough, the interstate-commerce hook would swallow the right to privately possess obscenity. At any rate, enforcing a ban on generating and possessing CG-CSAM for purely private, personal use is infeasible as a practical matter in a free society.

These are just a few hypothetical policy interventions that legislators might undertake; surely many other options exist. I raise them not to endorse or reject them but, rather, to illustrate how complex, careful, and nuanced the policy response to CG-CSAM will have to be.

CONCLUSION

The advent of technology for creating photorealistic computer-generated CSAM will worsen the existing issues caused by the online trade in photographic CSAM. Strategies for addressing this new variant on an old problem are complicated by the fact that under current First Amendment jurisprudence, some CG-CSAM images are protected speech whereas others are not. Policymakers must resist the temptation to propose unconstitutional legislation that would just get struck down in the courts. Instead, they should focus on measures that don't just uphold Americans' constitutional rights, but will meaningfully help American society—including children, online service providers, NCMEC, and law enforcement—to survive the coming flood.

The Digital Social Contract paper series is supported by funding from the John S. and James L. Knight Foundation and Meta, which played no role in the selection of the specific topics or authors and which played no editorial role in the individual papers.

---

[107] The statute prohibits possession of child obscenity that "has been shipped or transported in interstate or foreign commerce by any means, including by computer, or was produced using materials that have been mailed, or that have been shipped or transported in interstate or foreign commerce by any means, including by computer." 18 U.S.C. § 1466A(d)(4).

[108] *Circuit City Stores, Inc. v. Adams*, 532 U.S. 105, 115 (2001).