

# The Ethics of Customizable AI-generated Pornography

Jonathan Lang

McPherson Eye Research Institute

Center for Sleep and Consciousness Studies

University of Wisconsin-Madison

Aaron Yarmel

Center for Ethics and Human Values

Department of Philosophy

The Ohio State University<sup>1</sup>

## **0: Introduction**

Over the past two years, a thorny ethical landscape has begun to emerge in the wake of unprecedented developments in pornography. After bypassing Stable Diffusion's filters in 2022, users of the software created sexual content featuring realistic depictions of children.<sup>2</sup> In early 2023, a popular Twitch streamer was mired in controversy when he was caught producing deepfake porn of a fellow streamer without her consent.<sup>3</sup> A few months later, a South Korean man was sentenced to 2.5 years for creating illicit pornographic images with AI.<sup>4</sup> Soon after, it was discovered that New Jersey high school students were using AI to create and share "nudes" of female classmates, and in early 2024 another group of students at a Beverly Hills high school was caught doing the same.<sup>5,6</sup>

Such incidents have captured the attention of the press, the public, legislators, and even applied ethicists. Most of the critical attention thus far has been lavished on cases of 'deepfakes,' which are generally understood as depictions of victims that were created by training a generative AI program on images of these victims without their consent. Such discussions have rightfully centered on questions of consent, and on the ways that deepfake can harm the individuals whose faces are depicted in them (e.g., by disrespecting those individuals or by tarnishing their

---

<sup>1</sup> Both authors contributed equally to this work.

<sup>2</sup> <https://www.vice.com/en/article/xgygy4/stable-diffusion-stability-ai-nsfw-ai-generated-porn>

<sup>3</sup> <https://videogames.si.com/news/atric-deepfake-scandal-stepdown>

<sup>4</sup> <https://www.businessinsider.com/man-jailed-using-ai-create-sexual-images-children-south-korea-2023-9>

<sup>5</sup> <https://www.cbsnews.com/losangeles/news/probe-underway-at-socal-school-after-students-reportedly-created-nude-ai-generated-images-of-other-students/>

<sup>6</sup> <https://www.cbsnews.com/losangeles/news/probe-underway-at-socal-school-after-students-reportedly-created-nude-ai-generated-images-of-other-students/>

reputation).<sup>7</sup> As a result, the harms considered have mainly been the potential harms to the actual person facially deepfaked.

Yet deepfakes of this sort are only one instance of a broader category of *Customizable AI-Generated Pornography* (CAIP): a technology that gives ordinary individuals the means to digitally generate their own pornographic materials and customize them in virtually unlimited ways. The very same technology responsible for deepfakes can generate pornography depicting *non-existent* people and entities, and depictions of actual people can be generated without the use of any photographs of those same people as source materials for generative AI programs. In all instances of CAIP, individuals, whether real or fictional, can be depicted as performing the most disturbing and illegal of sexual acts, and at the highest degrees of photorealism. Moreover, the users themselves, as opposed to pornographic production companies, can direct the creation of such depictions at blindingly fast speeds and in complete accordance with their whims.

The nearly exclusive focus on a particular kind of deepfake has left us unprepared to grapple with the ethical and legal consequences of these other forms of CAIP. In the case of CAIP that depicts non-existent people, there is no actual individual who is depicted at all, and so questions about consent or harm to the individual depicted are inapposite. And while there is a growing movement to ban deepfake pornography that was produced by manipulating photographs and videos of actual people, these laws are poorly suited to regulating other methods of generating equally photorealistic depictions of the same people.

We here present a first attempt to address these lacunae. We begin by defining CAIP and showing how it is fundamentally different from other types of pornography that have preceded it. To this end, we identify five features of CAIP that render it a novel form of porn production (§I). We then argue that CAIP presents novel ethical problems that make it potentially far more harmful than traditional pornography. We begin with a discussion of machine learning, focusing on CAIP's ability to undermine people's agency due to its increased potential for addiction and the opportunity for corporations to use it for manipulative ends (§II).

We then turn to feminist arguments against pornography and show that most of the features that feminists are concerned about in traditional pornography are actually amplified in CAIP. In particular, CAIP presents new opportunities for sexual objectification, vastly expands the kinds and degrees of sexualized physical harm that can be depicted photorealistically, and allows men to sexualize photorealistic depictions of the very same women with whom they will interact in their lives (§III). Finally, we discuss the novel legal implications of CAIP. We focus on the challenge of adapting existing deepfake laws to apply to other forms of deepfakes that are produced without source material images of the victims before turning to the challenge of regulating the simple possession, without intent to distribute, of photorealistic, fictional depictions of child sexual abuse (§IV). We conclude with a call for others to continue the inquiry of which this paper is a first step (§V).

---

<sup>7</sup> <https://www.prindleinstitute.org/2024/02/the-wrong-of-explicit-simulated-depictions/>

## **I: What is CAIP and how is it different from other forms of pornography?**

*Customizable AI pornography* (CAIP) is any piece of media (i) whose content is pornographic<sup>8</sup>, (ii) that is created wholly or partially via an AI-generative process, and (iii) that allows the consumer to *participate* in that generative process in such a way that the consumer is able to dictate properties of the resulting pornography's depicta (and not just features of its playback). *How* the consumer may participate in the generative process can vary widely: she may feed the system an image of a real person's face and dictate that the AI graft it onto the head of an actor in an existing porn scene; she may use the AI to generate representations of *non-existent* humans engaged in pornographic acts; she may use a short text prompt to specify the general kind of act she desires to watch and then let the system generate a scene satisfying her basic prescription; she may deliver a wealth of information on her preferences to the system (about body type, scene location, types of acts, dialogue, etc.) to create a highly-specific scene tailored to her fetishes; or she may forgo any prompts and allow the AI system to develop a scene for her based on what the system has learned about her tastes from previous interactions.

CAIP is defined both in terms of the *process* used to create the pornography and the *control* that a consumer has over the products of that process. The process by which CAIP is produced is undoubtedly new, making it different from other types of porn. In addition, the *product* of these processes is also new: AI generative processes are capable of producing pornography with properties that are absent from the products of traditional methods of pornography production (legal or illegal). Finally, the sort of control afforded to consumers by CAIP is, we argue, also new in both type and degree. More specifically, in terms of its process, product, and consumer control, we argue that CAIP is unique in terms of having the following five properties:

### **(1) CAIP *can widely vary in the type and degree of indexicality that it exhibits.***

It is often argued that traditional digital and analogue methods of photography have an epistemic status different from other man-made representational artifacts<sup>9</sup>. This is because, unlike paintings and drawings, generating photographs entails mechanically capturing light from *real* objects on a photosensitive material. This process leaves behind an actual record of an existing place, object and/or event that occurred at the time at which the film emulsion or digital sensor was exposed. This characteristic casual provenance has come to be known as the *indexicality* of photography and film, and it is in virtue of this indexical bond to the actual world that photos and film license us to infer that their depicta actually exist (or, at the very least, existed).

Unlike true photographs, CAIP is able to represent places, events and individuals that do not exist at all. Such images that depict only non-existent things are *purely non-indexical* because no part of the image is technically a record of anything and does not refer to any extant person, object, or scene. Non-indexical CAIP can also *represent* things that do exist or have once existed, but without technically being an indexical record of anything that is being represented. For instance, imagine a 3D model of a real person's face and body generated from either a training

---

<sup>8</sup> There is a deep literature on how exactly to define pornography, but we will not engage in such debates here.

<sup>9</sup> "The objective nature of photography confers on it a quality of credibility absent from all other picture-making" ([Bazin, 1967], 14).

data set that houses actual photos of that person or merely from merely highly-descriptive text prompts that don't exploit any actual photos of that person in the training set. We call the former *matched deepfakes* and the latter *unmatched deepfakes*. In either case, such a model can be used to generate images of that person performing actions that they have never performed in places they have never visited. Such images we call *representative non-indexical*, and encompasses the category of deepfakes.

CAIP can also be *partially indexical* in that parts of the scenes it depicts are culled from actual photographs while other parts are AI generated (e.g., an AI system might generate an image of someone who has never existed and then implant it in the foreground of an actual photograph of a location, or the system might generate the face of a non-existent person and overlay it onto existing indexical pornographic footage). It can also use only indexical photographs to create a composite that, on the whole, is not indexical (e.g., a photo of bird wings composited onto a photo of a horse to create an image of a pegasus).

## **(2) CAIP can be *perceptually indistinguishable* from actual photographic media that exhibit true indexicality.**

At one point in history, a physical representation's being of great realistic detail (especially upon close inspection) supervened on its indexicality: If something looked like a mechanically produced, incredibly realistic depiction of a real scene then it was an indexical photograph. Yet as digital photography came to prominence and photo manipulation software became more powerful and ubiquitous, doctored images began to be perceptually indistinguishable from actual, unmanipulated photographs. Discerning a fully indexical photo of a scene from a composited image of other actual photographs' content (and perhaps altered further by painterly digital techniques) became exceedingly difficult. Still, it is arguable that during this period even digitally composited and altered photos laid claim to some level of indexicality, as it was difficult to produce fully photorealistic images (e.g., of a human face) without having to cannibalize portions of actual photographs for compositing source material. As for video, not even some of the best CGI studios could fool movie viewers into thinking that the images they saw onscreen were a fully-indexical: The technology was often too limited to render things like realistic motion, surface reflectance and textures (e.g., realistic skin on CGI characters)<sup>10</sup>.

With the advent of generative AI, all of this changed. AI systems can easily produce still images and video that are, to both the naked eye and to AI classifiers<sup>11</sup>, completely indistinguishable from an actual photographic record. In fact, in many instances these images are now erroneously judged by viewers to be even *more* indexical than actual photographs (a phenomenon known as "AI hyperrealism")<sup>12</sup>. Contemporary AI systems trained on data sets of millions of images can produce novel images of unparalleled fidelity and verisimilitude and with few limitations. The video output of such systems is becoming more realistic by the day, with millions being invested in AI-based production houses that promise to produce entire films without the use of cameras or

---

<sup>10</sup> <https://www.youtube.com/watch?v=NvFoKkWyZ5Y>

<sup>11</sup> <https://www.wsj.com/articles/ai-created-images-are-so-good-even-ai-has-trouble-spotting-some-8536e52c>

<sup>12</sup> Miller, E. J., Steward, B. A., Witkower, Z., Sutherland, C. A. M., Krumhuber, E. G., & Dawel, A. (2023). AI Hyperrealism: Why AI Faces Are Perceived as More Real Than Human Ones. *Psychological Science*, 0(0). <https://doi.org/10.1177/09567976231207095>

actual actors.<sup>13</sup> There has never been any process in existence that allows for the creation of nonindexical and partially indexical images that are absolutely perceptually indistinguishable from truly indexical photographs.

### **(3) The processes that produce CAIP are *automated and highly efficient***

Producing photorealistic pornography once required tools (e.g., cameras) and craft knowledge of how to use those tools. Time and effort were also required to recruit participants for scenes and to shoot and edit footage. On the pre-production side, scenes needed to be written or storyboarded. If the production was done in accordance with local and federal law, permits needed to be obtained and crews hired. Budgets and shooting schedules needed to be drafted, approved, and followed.

Generative AI is changing all of this. The automated and efficient nature of the AI image-generating systems mean that images and video can be produced at lightning speeds and require little effort or know-how on the part of the consumers using those systems. This, along with increased accessibility of these generative tools, has led to an explosion of AI generated images: According to some figures, as of 2023 there were already 15 *billion* AI images that had been created across the platforms Adobe Firefly, Dall-E 2, Stable Diffusion and Midjourney, a number which rivals the estimated number of photographs taken since the inception of photography.<sup>14</sup> The figures are surely higher than this, as the data reviewed only takes into account four of the many image generating apps that have become available.

The ease of use, efficiency, speed of production, and democratization of these AI tools have causally contributed to a proliferation of AI-generated images. It is reasonable to assume that this same causal relationship will hold more specifically in the case of CAIP. Further, individuals' exposure to novel images of illicit and illegal acts will surely increase, as AI systems allow people to easily manufacture such nonindexical images without (a) assuming the risk involved in procuring or producing indexical pornography of those same acts and (b) being deterred by the upsetting fact that actual people are harmed in the production. (Indeed, as we discuss below, there is already evidence that the amount of purely nonindexical CAIP and representative non-indexical CAIP depicting children is growing at a rapid rate.<sup>15</sup>)

### **(4) The set of potential depicta of CAIP is *virtually unbounded***

CAIP allows producers to create media that would not have otherwise been financially or logistically feasible, legally permissible, or consistent with basic facts about biology and the laws of physics. AI systems can effortlessly generate CAIP that depict creatures like aliens, monsters, animals, anthropomorphized objects and mutated humans. It can set its scenes in exotic locations like space or in real locations that would be inaccessible for real life film shoots due to cost or inability to obtain permission. It can depict sexual crimes involving adults and minors. It can feature the likenesses of real individuals that would never willingly consent to performing the acts depicted.

---

<sup>13</sup> <https://variety.com/vip/how-generative-ai-could-enable-a-new-era-filmmaking-1235898355/>

<sup>14</sup> <https://journal.everypixel.com/ai-image-statistics#:~:text=In%20total%2C%20more%20than%2015,Midjourney%2C%20and%20DALLE%2D2.>

<sup>15</sup> <https://www.theguardian.com/technology/2023/oct/25/ai-created-child-sexual-abuse-images-threaten-overwhelm-internet>

In terms of its nearly limitless abilities, CAIP resembles the media of pornographic drawings or cartoons. But while pornographic drawings have existed since antiquity and pornographic cartoons since the late 1920's<sup>16</sup>, never has there been a method of creating pornography that can rapidly and automatically produce these depictions in such a way that they are indistinguishable from indexical photographs and video. In other words, never has there been a way to produce pornography that exhibits both the described unboundedness and the abovementioned properties (2) and (3).

**(5) CAIP can be generated and customized by users to a degree that was hitherto impossible.**

Even before the advent of CAIP, consumers have long evinced at least some control over the pornography that they consumed. Consumers have been able to select specific materials they wanted to view from a diverse pool of available pornographic materials. The explosion of digital porn in the internet era has provided consumers with more choice than ever, with powerful search engines allowing consumers to easily find the highly-specific types of pornography that they desired to view. Consumer preferences monitored by the porn industry are used to identify preference trends and used to shape the content of new porn that is produced<sup>17,18</sup>. Finally, consumers not part of the porn industry proper have sometimes used home recording equipment to produce their *own* amateur pornography featuring themselves and partners. Such individuals have certainly been able to exhibit control over the porn-generating process and its products in a way similar to individuals using AI.

Still, never have consumers had the sort of control over pornographic content as the control promised to them by CAIP. This new type of pornography makes every consumer a creator, with full authorial control over the resulting product and its properties. No industry intermediaries stand between the process of creation and consumption. No actors, recording equipment, or know-how of image or video production are required. Text prompts can now effortlessly generate any desired image, and every possible parameter of the representational outputs of the generative systems seem up for adjustment by the consumers that use them.

Other types of pornographic media may possess some proper subset of the above five features, but only CAIP has all five. Yet while this shows that CAIP is different from the pornographic media that has preceded it (in terms of process, product and consumer control), it does not necessarily show that this novelty generates new and interesting ethical concerns. We now turn our attention to the question of how these properties of CAIP force us to acknowledge that CAIP is of even more concern than traditional forms of pornography. More specifically, in what follows we argue that the unique nature of CAIP *complicates* extant debates on porn, *amplifies* pornography's already recognized morally problematic effects, and *generates* new ethical and legal issues pertaining to the production, distribution, and consumption of pornography. We begin by showing that this technology produces a series of new ethical issues that have never yet arisen in the history of pornography.

---

<sup>16</sup> See Cohen, Karl F. *Forbidden animation: Censored cartoons and blacklisted animators in America*. McFarland, 2013. Eveready Harton in *Buried Treasure* (1929)

<sup>17</sup><https://qz.com/1407235/porn-sites-collect-more-user-data-than-netflix-or-hulu-this-is-what-they-do-with-it>

<sup>18</sup> Keilty, Patrick. "Desire by design: pornography as technology industry." *Porn Studies* 5.3 (2018): 338-342.

## **II. CAIP, agency, and algorithms**

It is no secret that social media platforms collect user data and deploy algorithms in an effort to determine what its users want to see. Such algorithms generate hypotheses about our preferred content, which are then tested by presenting users with that content and seeing whether that content increases engagement and time spent on the platform. Even after the algorithms bolster our engagement, they will continue to learn about us, experiment with us, and refine their predictions, all in an effort to capture and keep our attention.

Similar to social media, porn platforms also heavily monitor their users' data and deploy algorithms to better understand their users' sexual preferences.<sup>19</sup> The purpose is ultimately the same as with social media: To better understand what content users want to see in order to get them to that content faster, increase their engagement with the platform, capture their attention, and uncover viewership trends that will help the platform understand what type of new content should be created and made available.

These practices are troubling given that pornography can be addictive<sup>20</sup>. Data mining and algorithms work in concert to shove into users' view existing content from which they find it hard to turn away, arguably enhancing the addictive quality of the media by manipulating the way it is presented to users. But notice that, even in this case, the *content* of each video itself is already fixed and is not being *generated* to exactly match the deepest desires of the user. Yet with CAIP, the situation is not the same. As was stated earlier, CAIP can be created quickly and automatically, is practically unbounded in what it can depict, and every aspect of it can be fully customized to meet the needs of its creator (properties 3, 4, and 5). Unlike the traditional pornographic content that populates porn platforms, CAIP can be *generated rapidly and even as a user is scrolling through a platform*. Every parameter of CAIP can be tweaked, allowing for hyper-specific hypotheses about user's desires to be deployed and tested on individual users. Never before in history has there been a time when such content could be generated on the fly to directly test and/or cater to the desire of each consumer.

When porn platforms can create their content from scratch with AI systems and every image parameter can be automatically and instantly adjusted by an algorithm, new and insidious opportunities arise for creating hyper-addictive pornography that can be fine-tuned to individual consumers' particular desires. CAIP-generating systems are thus a powerful tool for *enhancing* pornography's addictive nature. Further, it is undeniable that porn platforms seeking to capture our attention have great financial incentive to increase the addictive nature of porn in just this way. Given this, CAIP's unique nature poses a real and novel threat to consumers' agency in a way that is absent with regard to traditional pornography.

---

<sup>19</sup> Keilty, Patrick. "Desire by design: pornography as technology industry." *Porn Studies* 5.3 (2018): 338-342.

<sup>20</sup> Love, Todd, et al. "Neuroscience of internet pornography addiction: A review and update." *Behavioral sciences* 5.3 (2015): 388-433.

We now know that the neural processes underlying porn addiction are the same basic mechanisms that underlying substance addiction.<sup>21</sup> Given this, to couple CAIP-producing systems with machine-learning algorithms in the way described above would be like intentionally increasing the addictive properties of a drug (e.g., an opioid) by using individuals' idiosyncratic neurobiological data to optimize its addictive effects for *each individual's particular neurobiology*. To the extent that one thinks that such a practice should be impermissible, so too should one think it impermissible to bolster the addictive nature of porn by tailoring its content so perfectly to a consumer's deepest desires. Hence, if pornography is harmful because of its addictive nature and corrosive effects on our agency, the potential for harm of CAIP becomes orders of magnitude greater, at least when image-generating systems can use machine learning algorithms to efficiently tap into our most powerful desires and use them against us.

When it comes to CAIP's deleterious effects on our agency, there is yet another worry. This has to do with CAIP's potential to modify our desires and our behaviors. Consider again how social media platforms are leveraged by corporate interests to capture our attention and maximize profit<sup>22</sup>. For these companies, a main value of capturing user attention is that it allows them to sell users products, show users paid advertisements, generate revenue through users' affiliate link clicks, etc. But how long will it be before the parameters being adjusted in the CAIP we see are used to try to sell us products or modify our behaviors? This technology promises to invade the most private and primal areas of our psyche, and there is a real danger that it may then be able to alter our desires in pernicious ways and exploit our psychological vulnerabilities.

### **III: CAIP Amplifies Anti-Porn Feminist Concerns**

Feminists disagree about the ethics of creating, consuming, and regulating pornography.<sup>23</sup> We take no stance in this debate, but instead argue for a conditional claim: feminists who are concerned about traditional pornography should be even more concerned about CAIP. This is because CAIP has an amplifying effect on most of the features of traditional pornography that feminist critics of pornography have linked to harms and wrongs against women.

Following A.W. Eaton, we refer to feminist arguments against pornography as Anti-Porn Feminist (APF) arguments and proponents of such arguments as Anti-Porn Feminists (APFs).<sup>24</sup> Some APFs define "pornography" in sexist terms: "material (pictures and/or words) that depicts women's subordination in such a way as to endorse, encourage or eroticize it."<sup>25</sup> In this section,

---

<sup>21</sup> Love, Todd, et al. "Neuroscience of internet pornography addiction: A review and update." *Behavioral sciences* 5.3 (2015): 388-433.

<sup>22</sup> Bhargava, Vikram R., and Manuel Velasquez. "Ethics of the attention economy: The problem of social media addiction." *Business Ethics Quarterly* 31.3 (2021): 321-359.

<sup>23</sup> West, Caroline, "Pornography and Censorship", *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition), Edward N. Zalta & Uri Nodelman (eds.), URL = <<https://plato.stanford.edu/archives/win2022/entries/pornography-censorship/>>.

<sup>24</sup> Eaton, A.W. (2007) "A Sensible Antiporn Feminism," *Ethics*: 117.

<sup>25</sup> West. See also Longino, H., 1980, "Pornography, Oppression, and Freedom: A Closer Look," in Laura Lederer (ed.), *Take Back The Night*, New York: William Morrow; MacKinnon, C., 1987, "Not a Moral Issue" and "Francis

we will restrict our focus to pornography with these inegalitarian features in order to make a case that is more broadly acceptable to feminists who may support so-called “feminist porn.”<sup>26</sup> For clarity, we will refer to such inegalitarian porn as *ipornography*.

Our claim is that CAIP has an amplifying effect in *most*, rather than all cases, and this is because some APFs base their arguments on how ipornography is traditionally *produced*.<sup>27</sup> Amanda Cawston, quoting Gail Dines, notes that pornographic actors are prone to experiencing “HIV; rectal and throat gonorrhea; tearing of the throat, vagina and anus; and chlamydia of the eye.”<sup>28</sup> Our view is that harms of this kind will most likely be decreased by CAIP, as CAIP can be produced without human performers.

In what follows, we will discuss a representative sample of APF arguments. We will then describe how CAIP amplifies the features that motivate these arguments.

### **(1) APF Arguments**

One style of APF argument posits a causal connection between the consumption of ipornography and the commission of acts of violence against women: “By eroticising rape, torture, cruelty and the objectification of women, pornography undermines prohibitions against rape and normalises violence and abuse, ultimately leading to increased rates of rape and assault.”<sup>29</sup> Catherine MacKinnon and Andrea Dworkin endorsed something like this view in their famous anti-pornography ordinances, which state that “any woman, man, child, or transsexual who is assaulted, physically attacked or injured in a way that is directly caused by specific pornography has a claim for damages against the perpetrator, the maker(s), distributor(s), seller(s), and/or exhibitor(s), and for an injunction against the specific pornography’s further exhibition, distribution, or sale.”<sup>30</sup>

Although the ordinance just quoted references a direct, causal relationship between viewing pornography and committing an assault, APF arguments are typically more nuanced. For example, Eaton, advances a probabilistic account of causation, such that consuming pornography makes it more likely that someone’s attitudes will be shaped in sexist ways, and these attitude

---

Biddle’s Sister: Pornography, Civil Rights and Speech,” in *Feminism Unmodified*, Cambridge, MA: Harvard University Press, pp. 146–162, 163–197.

<sup>26</sup> For an overview, see Taormino, T., Shimizu, C.P., Penley, C., & Miller-Young, M. (2013). *The Feminist Porn Book: The Politics of Producing Pleasure*. New York: The Feminist Press.

<sup>27</sup> MacKinnon, Catharine. 1996. *Only Words*. Cambridge: Harvard University Press. MacKinnon, Catharine. 2005. *X Underrated*. *Times Education Supplement*. London, May 20, 2005. Farley, Melissa. 2006. “Prostitution, Trafficking, and Cultural Amnesia: What We Must Not Know in Order to Keep the Business of Sexual Exploitation Running Smoothly.” *Yale Journal of Law and Feminism*, 18: 101–136.

<sup>28</sup> Amanda Cawston (2019) *The feminist case against pornography: a review and re-evaluation*, *Inquiry*, 62:6, 628. Dines, Gail. 2012. *Porn, Syphilis, and the Politics of the Money Shot*. *The Guardian*, August 28. <http://www.theguardian.com/commentisfree/2012/aug/28/porn-syphilis-money-shot-condoms>.

<sup>29</sup> Cawston, 631.

<sup>30</sup> (1985) “Appendix: The MacKinnon/Dworkin Pornography Ordinance,” *William Mitchell Law Review*: Vol. 11: Iss. 1, Article 5. Available at: <http://open.mitchellhamline.edu/wmlr/vol11/iss1/5>, 123.

changes will then reinforce inequalities between men and women.<sup>31</sup> She makes her case by first describing a set of background conditions: our society contains inequalities between men and women, such that women occupy a socially subordinate position, and these inequalities are gravely unjust; these inequalities are “sustained and reproduced” by the socialization of children into gender roles and through violence and force (such as sexual assault); many people find that some aspects of gender inequality--such as dominance/strength vs softness/submissiveness and even sexual violence--are erotically appealing. When juxtaposed against these background conditions, the harm of pornography is as follows: “by harnessing representations of women’s subordination to a ubiquitous and weighty pleasure, [ipornography] is especially effective at getting its audience to internalize its inegalitarian views.”<sup>32</sup>

Other APFs offer detailed accounts of how pornography shapes our attitudes towards sex. According to Joan Mason-Grant, the existence of sexual desire is natural, but “it takes shape within the social context in which we come to maturity as sexual beings.”<sup>33</sup> This is analogous to how hunger is a natural desire, but our hunger for a specific food--beans and rice, meat, fresh fruit, potato chips, oreo cookies--is shaped by our experiences with it. As people consume ipornography and engage in sexual behaviors while doing so, the shape of this natural sexual desire is molded by those experiences. Drawing from Andrea Dworkin, Mason-Grant describes this shaping process as a sexual education in which viewers “learn in their bodies about women from the pornography in a way that it doesn’t matter what they think.”<sup>34</sup> Through this learning process, viewers come to treat their learned desire for the sexual subordination of women as if they were natural sexual desires given to them by nature. Moreover, the self-centered nature of the activity trains viewers to objectify women: they experience a “cycle of erotic desire, arousal, and satisfaction [that is] organized arrogantly, with exclusive concern for the needs, desires, and involvement of the user.”<sup>35</sup>

Another major line of APF argumentation is that ipornography silences women by creating a social climate in which three phenomena are observed: women are reluctant to speak at all; when women do speak, their opinions are ignored when they contradict “the picture of women contained in pornography”; and when women speak, their speech is misunderstood because their words are interpreted based on the expectations set by pornography.<sup>36</sup> When a woman says “no” to a man who is attempting to initiate sexual contact with her, he might ignore what she is saying (the second feature) or reinterpret her “no” as an act of flirting in the service of increasing arousal (the third phenomenon). And if he sexually assaults her, she may be reluctant to press criminal charges afterwards due to the hostile and adversarial environment into which she would be placed (the first phenomenon).

---

<sup>31</sup> Eaton, 680, 697. Eaton defends an account of probabilistic causation: “x is a cause of y if and only if (i) x occurs earlier than y and (ii) the probability of the occurrence of y is greater, given the occurrence of x, than the probability of the occurrence of y given not-x” (697).

<sup>32</sup> *ibid.* 679-680.

<sup>33</sup> Mason-Grant, J (2013). “Pornography as Embodied Practice,” in *The Philosophy of Sex*, sixth edition, Nicholas Powers, Raja Halwani, and Alan Soble. Rowman and Littlefield Publishers, Inc., Lanham, MD, USA., 532.

<sup>34</sup> *ibid.* 533.

<sup>35</sup> *Ibid.* 534.

<sup>36</sup> West.

As was the case for the harm-based arguments discussed above, APFs have offered nuanced silencing arguments that include analyses of the silencing itself and accounts of the harms and wrongs inflicted by it. Jennifer Hornsby and Rae Langton,<sup>37</sup> drawing on earlier work by MacKinnon, have developed an account of the ways women are wronged by the silencing based on a right to freedom of speech: women are deprived of the benefits guaranteed by a right to freedom of speech, and defending women's rights to freedom of speech requires restrictions against such deprivations. And Ishani Maitra has offered a novel analysis of the phenomenon of silencing itself based on Paul Grice's work on communicative intent.<sup>38</sup>

## (2) CAIP's Amplification Effects

Regina Rini and Leah Cohen (2022) have focused on novel harms associated with matched deepfakes: a "digitally manipulated face of one woman [is] stitched onto the body of another."<sup>39</sup> They argue that this kind of ipornography, which they call "frankenporn," introduces a new form of objectification. Instead of a performer who has thoughts, values, and feelings, such CAIP is just a "composite of the parts of two people."<sup>40</sup> Their concern does not appear to be merely that the consumer is viewing an image of a person who does not exist, nor that he is viewing an image of a real individual. Their concern is that the consumer is *combining* these two practices: he thinks of an image constructed from combinations of features as if it were a representation of an actual person, but he also knows that the images he is consuming are not of someone with desires, beliefs, or a point of view: "the entity depicted in frankenporn *cannot* have a determinate will, since it is the composite of the parts of two different people, unified only in digital artifice."<sup>41</sup>

While Rini and Cohen do not make this point explicitly, an APF should be concerned that such a consumer will train himself to become sexually aroused by a fabricated version of a real person that has been stripped of agency, intentionality, and desires. He will then come to expect that actual women, including the one he is imagining, will be limited in this way. A version of this APF argument will still apply even if, contra Rini and Cohen, viewers of frankenporn attribute rich inner worlds to the depictions they are consuming in a manner analogous to a reader's empathizing with a favorite character from a fiction novel. Such viewers may imagine women who always desire abuse, subjugation, and sexual contact and only pretend to refuse sex in an

<sup>37</sup> Langton, Rae & Hornsby, Jennifer (1998). Free speech and illocution. *Legal Theory* 4 (1):21-37.

<sup>38</sup> Maitra, I., 2009, "Silencing Speech," *Canadian Journal of Philosophy*, 39(2): 309–338. On Maitra's account, silencing is understood as *communicative disablement*. What it means for a woman to communicate that she is refusing to have sex is for her to intend the following three things:

- (i) her audience come to believe that (all things considered) she does not want to have sex with him;
- (ii) her audience come to think (recognize) that she has the informative intention mentioned in (i); and,
- (iii) her audience's fulfillment of (ii) give him a reason to fulfill (i) (326).

The phenomenon of silencing amounts to a failure of (ii) or (iii). For example, an audience may fail to notice or correctly interpret what she is saying, or the audience may not view her desire for the sex to not take place as a reason to stop.

<sup>39</sup> Rini, R., and Cohen, L. (2022) "Deepfakes, Deep Harms" *Journal of Ethics and Social Philosophy*, 22:2. 146.

<sup>40</sup> *ibid.* 147.

<sup>41</sup> *ibid.* 147.

effort to increase sexual tension. These concerns apply to unmatched deepfakes as well, as impoverished or inegalitarian psychologies can also be attributed to depictions generated in response to highly specific text prompts.

Due to CAIP's unboundedness, indistinguishability, and efficiency, photorealistic, erotic depictions of torture, rape, and cruelty that are biologically impossible or would result in great bodily injury or death can be generated extremely quickly. Moreover, as we have noted, the customizability of CAIP means that the stars of such depictions may be one's peers: neighbors, classmates, colleagues, coworkers, and friends. We refer to such CAIP as *peer-porn*.

In the APF arguments surveyed above, there are several steps between watching ipornography and performing behaviors that harm or wrong women: a man consumes ipornography, which causes him to change his attitudes, which leads to behavioral changes that culminate in rights violations, the maintenance of oppressive social practices, and other harms to women. But a man shaped by peer-porn is engaging directly, and from the start, with depictions of the very same women he encounters in his life. As we have noted, these depictions may be unprecedented in their capacities for objectification and photorealistic violence. And given its efficiency, CAIP promises a world where, for some men, the amount of time spent interacting in person with a specific woman in one's life is dwarfed by the amount of time spent consuming ipornographic depictions of her. While fantasizing about one's peers is not novel, the CAIP in question involves photorealistic representations rather than degraded images in the mind's eye.

So far, we have just discussed deepfakes, but recall that there also exists pure non-indexical CAIP. The APF concern here is twofold. First, consumers may be more likely to consume such ipornography because they can assuage their guilty consciences by reminding themselves that no actual women are depicted in it. However, there is no reason to think that pure non-indexical CAIP is any less likely to train the men who consume it to enjoy sexualized depictions of the subordination of women. What's more, one might worry that consuming such CAIP would amplify Rini and Cohen's concerns far beyond the frankenporn that they discuss, for there is not even a real woman whom a consumer is *attempting* to represent.

#### **IV: CAIP has Novel Legal Implications**

Different forms of CAIP generate distinct and novel legal implications, which we will discuss in this section. The first implications concern deepfakes. Existing and proposed deepfake regulations are often written in such a way that they do not apply to all representative non-indexical CAIP, but only to what we called *matched deepfakes*. The motivation to ban deepfakes apply equally well to what we called *unmatched deepfakes*, but we will demonstrate that attempts to ban them run the risk of banning forms of expression that should be protected.

The second implications concern a particularly controversial manifestation of pure non-indexical CAIP: pornographic fictional depictions of child sexual abuse. As we will show there are good reasons, based on the harm principle, to be concerned about the simple possession of such CAIP even in cases where there is no distribution, receipt, solicitation, or intent to distribute, receive,

or solicit such materials. However, a restriction against the simple possession of such fantasy materials is likely to be unconstitutional.

### **(1) regulating representative non-indexical CAIP**

In the United States, there are currently no federal laws that ban or regulate deepfake pornography *per se* (i.e., as opposed to regulations on other classes of conduct under which cases of deepfake pornography may fall). However, many states have passed regulations that address deepfakes in particular, and there have been several attempts at federal regulation. The challenge is that all existing or proposed regulations either go too far in the direction of restricting speech or else fail to ban unmatched deepfake CAIP that is perceptually indistinguishable from unmatched deepfake CAIP.

Consider three peer-porn vignettes in which Jennifer creates and then texts a sexually explicit depiction of her crush, Enrique, to her friends Akira and Stephanie. In all three cases, Jennifer does not seek to gain commercially from the production of the depiction, Enrique's reputation is unaffected because Akira and Stephanie do not think that the picture is a representation of real events, and all parties are adults:

*A: Jennifer draws a naked picture on her iPad that resembles her crush, Enrique.*

*B: Jennifer uses a computer program on her iPad that turns text prompts into images. After an hour of experimenting, she is able to cajole the program into outputting a picture of a naked man that resembles her crush, Enrique.*

*C: Jennifer trains a learning algorithm on a set of pictures of her crush, Enrique. After inputting every picture and video she is able to find, she uses the program to generate a depiction of a naked man that resembles Enrique.*

For each vignette, we can imagine two permutations: in permutation 1, the image created by Jennifer is a photorealistic depiction of Enrique; in permutation 2, it is not. Before discussing which of Jennifer's actions ought to be illegal, we will begin by exploring which activities *are* or *would be* illegal under existing or proposed regulations.

In January of 2024, the No Artificial Intelligence Fake Replicas and Unauthorized Duplications (No AI FRAUD) Act was introduced into the United States Congress by María Elvira Salazar (R-FL) and Madeleine Dean (D-PA).<sup>42</sup> If passed, one consequence of this bill will be effectively a ban on publishing (construed broadly) all sexual digital depictions of a living person unless this person (or others to whom this right was transferred) consented to the creation of said depictions. The draft text defines “digital depiction,” as “a replica, imitation, or approximation of the likeness of an individual that is created or altered in whole or in part using digital technology,” thus including within its scope an action as innocuous as drawing a sexually explicit cartoon of adult characters that are meant to resemble actual people. Similarly, California's Assembly Bill No. 602 CHAPTER 491 criminalizes the creation and disclosure, or even the mere disclosure, of

---

<sup>42</sup> <https://www.congress.gov/bill/118th-congress/house-bill/6943?s=1&r=9>

a sexually explicit depiction of someone without that person's consent.<sup>43</sup> Both bills would, thus, include all versions of all three vignettes even if the image of Enrique is not photorealistic.

On the other hand, Texas's deepfake law is narrower. It defines a deepfake as follows: "Deep fake pornography is when the likeness of one person is digitally altered to look like someone else and is inserted into a video in which sexual conduct is occurring or intimate parts are exposed."<sup>44</sup> In unmatched deepfakes, the likeness of one person is not digitally altered because the source material does not include any indexical images that depict the the individual depicted in the CAIP. The same feature can be found in Illinois's Public Act 103-0294,<sup>45</sup> New York's S1042A,<sup>46</sup> and Virginia's 18.2-386.2.<sup>47</sup> For this reason, both permutations of Vignettes A and B would be allowed.

Minnesota's deepfake ban is both wide and narrow. It is narrow because the definition of "deepfake" applies to a "video recording, motion-picture film, sound recording, electronic image, or photograph, or any technological representation of speech or conduct" that is "so realistic that a reasonable person would believe it depicts speech or conduct of an individual," and it is wide because it merely specifies that the production of the deepfake "was substantially dependent upon technical means, rather than the ability of another individual to physically or verbally impersonate such individual."<sup>48</sup> An individual may be humiliated by a video appearing to show them performing sexual activities even if the creators of the video are careful to include signals that point to nonindexicality (e.g., bizarre artifacts, a banner across the screen that indicates that the video does not depict real events, or even the presence of extraterrestrial organisms). And an electronic image created by an individual with a stylus and a drawing pad could, in principle, be so photorealistic as to confuse a reasonable person. Thus, some cases of Vignette A will be banned and some cases of Vignette C will be allowed.

The challenge of regulating deepfakes is, thus, as follows: one can either regulate images based on the processes by which they were created or based on their visually perceivable features. The former option leads to regulations on matched deepfakes, but not perceptually indistinguishable unmatched deepfakes. This is a problem because the justification for banning deepfakes is typically based on the dignitary, reputational, and other harms inflicted upon the individuals who are depicted, the contribution of the practice to sexist norms (as we discussed above), and the effects of misinformation when viewers believe that the images depict real events.<sup>49</sup> Since there is no reason to distinguish between matched and unmatched deepfakes on any of these grounds, the former option leads to regulations that are too narrow.

If one chooses the latter option, one will also run the risk of banning what ought to be protected forms of expression. Intuitively, it would be bizarre if it were illegal to share with one's friends, for no commercial gain, a hand-drawn image that depicts a naked adult in a context where the

---

<sup>43</sup> [https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill\\_id=201920200AB602](https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB602)

<sup>44</sup> <https://capitol.texas.gov/tlodocs/88R/analysis/html/SB01361F.htm>

<sup>45</sup> <https://www.ilga.gov/legislation/publicacts/fulltext.asp?Name=103-0294>

<sup>46</sup> <https://www.nysenate.gov/legislation/bills/2023/S1042/amendment/A>

<sup>47</sup> <https://law.lis.virginia.gov/vacode/title18.2/chapter8/section18.2-386.2/>

<sup>48</sup> [https://www.revisor.mn.gov/bills/text.php?number=HF1370&type=bill&version=3&session=ls93&session\\_year=2023&session\\_number=0](https://www.revisor.mn.gov/bills/text.php?number=HF1370&type=bill&version=3&session=ls93&session_year=2023&session_number=0)

<sup>49</sup> Rini and Cohen (2022).

image will not be wrongly viewed as indexical. This is due to the intuition that the state should not intervene so deeply into such semi-private instances of expression.<sup>50</sup>

Our position is not that successful regulation is impossible, but merely that CAIP raises novel legal puzzles that cannot be solved by the same tools developed to deal with traditional deepfakes. Such regulations are inadequate to strike the right balance between protecting our rights to freedom of expression and protecting the dignity of victims in this new reality.

## **(2) the harder challenge of regulating pure non-indexical CAIP**

While regulating deepfakes is challenging, things are even more difficult in the case of pure non-indexical CAIP. We will focus on a particularly salient example, which is fictional depictions of child sexual abuse (FDCSA).

The harm principle is commonly appealed to in arguments for the restriction of pornography. Hadeel Al-Alosi expresses this principle as, “state interventions may be justified to prevent harm to others, but otherwise freedom takes priority.”<sup>51</sup> We will sidestep the rich literature about finding the right balance between protecting freedoms and minimizing harms here because it is uncontroversial, on any account, that sexual abuse constitutes an extremely serious harm. For this reason, we will assume that all reasonable interpretations of the harm principle will hold that child sexual abuse should be illegal, and, thus, that creating an indexical pornographic depiction of child sexual abuse should also be illegal. We will also grant, for the sake of argument, that it is legitimate for states to intervene to prevent indexical and representative non-indexical versions of FDCSA because of the risk that the particular children being depicted will suffer dignitary harms if they ever discover the uses to which their likenesses have been put.<sup>52</sup>

It is also plausible that some interventions to ban the *publication* of pure non-indexical CAIP are legitimate based on the harm principle.<sup>53</sup> From September 1st through September 31st of 2023, the Internet Watch Foundation investigated a forum on the dark web that focuses on child sexual abuse materials. During that month, they found that 2,978 depictions of child sexual abuse were posted, many of which were realistic enough to be indistinguishable from actual images of

---

<sup>50</sup> Another consideration is that the harms inflicted by deepfakes emerge in a context where some people believe that they are records of actual events. It is possible that deepfakes will come to saturate the internet to such an extent that *nobody* will believe that the sexually explicit images they encounter on the internet will be documentary, thus depriving deepfakes of their power to cause these harms. We are unsure how plausible this is.

<sup>51</sup> Al-Alosi, H. (2018). *The criminalisation of fantasy material: law and sexually explicit representations of fictional children*. New York, NY: Routledge. 84.

<sup>52</sup> We do not claim that *all* state interventions undertaken for the purpose of preventing sexual abuse of children are justified by the harm principle. It would surely be impermissible to force all adult members of a population to undergo a series of invasive physiological examinations designed to determine whether they could ever be sexually aroused by representations of children and then euthanize anyone who receives a positive test result. It also does not follow that any tool that could be possibly be used to facilitate sexual abuse should be restricted. As Al-Alosi notes, such a principle could “lead to the criminalization of various items that may conceivably be misused by child molesters...” such as “...candy and children’s toys...” (94). A ban on such items would go too far, and for the same reason that it would go too far to ban pens and pencils on the grounds that they can be used to create acts of libel.

<sup>53</sup> Another argument could be grounded in the offense principle, but we will not devote the resources necessary to an exposition of this principle here because the harm principle, and not the offense principle, is relevant to the salient example in this section: the simple possession of pure non-indexical CAIP.

children. Due to the perceptual indistinguishability and efficient production of such CAIP, we have already reached a point where law enforcement officers cannot reliably distinguish between visual records of actual child abuse and fantasy materials.<sup>54</sup> Neil Levy's observation from 22 years ago is, thus, no longer true: "given today's technology, it is not very difficult for experts to distinguish between virtual and actual pictures [of child pornography]."<sup>55</sup> Such confusion results in costly errors in two directions: fantasy materials are mistaken for actual records of child abuse, and actual records for fantasy materials. In the former case, false leads waste valuable law enforcement resources; in the latter case, real abuse is overlooked. Either way, the result is a reduction in the number of real victims found and saved, as well as offenders arrested and prosecuted.

The most challenging cases involve the simple possession of pure non-indexical CAIP that is perceptually indistinguishable from indexical depictions of child sexual abuse. To motivate this challenge, consider the following series of fictional vignettes, all of which take place within the privacy of the home of a man named Dennis:

- A. Dennis draws a sexually explicit image of a fictional child on a piece of paper, which he then looks at and subsequently incinerates in his fireplace.
- B. Dennis draws a sexually explicit image of a fictional child on a computer not connected to the internet on a drawing tablet with a stylus. He looks at the image, and then erases the file completely from his computer.
- C. Dennis draws a sexually explicit image of a fictional child on a computer not connected to the internet on a drawing tablet with a stylus. This computer has a generative AI program on it, and he uses this program to modify the features of the fictional child to suit his preferences. He then looks at the image and subsequently erases the file completely from his computer.
- D. Dennis has a computer with a generative AI program on it that is not connected to the internet. He requests that this program create a sexually explicit image of a fictional child. He then looks at this image and subsequently erases the file completely from his computer.

Our question is, has Dennis performed an action, in any of these cases, that it would be legitimate for a state to punish?

As a matter of descriptive fact, different countries have answered this question differently. Al-Alosi notes that Australia and the United Kingdom ban the simple possession of fictional child pornography and "make no exception for privately kept, self-produced fantasy material."<sup>56</sup> This restriction applies even to cartoonish depictions of fictional characters, as was the case in *McEwan v Simmons*. In this 2008 case, a man was convicted of possessing and of accessing child pornography because he downloaded "a series of cartoons depicting the child characters from the television program, *The Simpsons*, engaging in sexual acts."<sup>57</sup> In 2013, another offender in Australia was convicted of possessing child pornography that also consisted of sexually explicit

---

<sup>54</sup> Internet Watch Foundation, 2023,

<https://www.iwf.org.uk/about-us/why-we-exist/our-research/how-ai-is-being-abused-to-create-child-sexual-abuse-imagery/>

<sup>55</sup> Levy, Neil (2002). Virtual child pornography: The eroticization of inequality. *Ethics and Information Technology* 4 (4):319-323. 320.

<sup>56</sup> Al-Alosi, 113.

<sup>57</sup> *ibid.* 37.

*The Simpsons-inspired* cartoons.<sup>58</sup> Thus, Dennis's conduct is illegal in Australia and the United Kingdom in all four vignettes—even if his depictions are no more realistic than cartoons of Bart and Lisa Simpson.

In the United States, it is most likely the case that none of these vignettes would be banned by existing laws. In each case, Dennis has not merely acted without an intent to distribute the depictions, but even taken steps to preclude that possibility (i.e., by ensuring that his computer is disconnected from the internet and destroying the images after viewing them). While it is constitutional to restrict obscene speech, the precedent set by the Supreme Court in the case *Stanley v. Georgia* is that obscenity regulations cannot ban the private possession of obscene materials in one's own home.<sup>59</sup> A ban based on the harm principle would also be inconsistent with past Supreme Court decisions. In particular, the Child Pornography Prevention Act of 1996 was found to be unconstitutional in the Court's decision in *Ashcroft v. Free Speech Coalition*. As Ratner notes, the Court found that virtual child pornography is not intrinsically related to child abuse, and that "the prospect of a crime... by itself does not justify laws suppressing protected speech."<sup>60</sup>

There has never been, nor (for obvious ethical reasons) is there ever likely to be, a controlled experiment capable of establishing a causal relationship between viewing FDSCA and committing acts of sexual assault against children. For this reason, an argument for criminalizing Dennis's conduct grounded in the harm principle must appeal to evidence that is suggestive of a causal relationship rather than conclusive. One source of evidence concerns the relationship between fantasizing about child sexual assault and committing actual acts of sexual assault. As Al-Alosi notes, there is "little dispute that there is a link between deviant sexual fantasies and childhood sexual abuse."<sup>61</sup> That said, the *nature* of this link is disputed: some evidence suggests that fantasizing about sexual abuse makes the fantasizer more likely to commit actual abuse, while other researchers "have also suggested that deviant fantasies may allow paedophiles to release sexual tension, which reduces the chances of paedophiles engaging in sex offending in real life."<sup>62</sup> Moreover, it is unclear that the fact that someone engages in sexual fantasies about children is *predictive* of the commission of acts of child sexual abuse, as "there is a substantial amount of research suggesting that fantasies indicative of paedophilia are prevalent amongst 'normal' people who are assumed to have never committed sexual abuse."<sup>63</sup>

While fantasizing about child sexual abuse has not been conclusively linked to abusing children, a stronger case can be made for a relationship between abusing children and viewing pornography that depicts actual child sexual assault. Let's begin, though, with a few reasons to doubt such a causal relationship. Levy has correctly noted that "the question whether ('ordinary') pornography encourages or causes rape has been debated for forty years now, yet the empirical evidence on both sides of the question seems indecisive."<sup>64</sup> This is to say, we do not know

---

<sup>58</sup> *ibid.* 37.

<sup>59</sup> Ratner, C. (2021) When "Sweetie" is not so Sweet: Artificial Intelligence and its Implications for Child Pornography, *Family Court Review*, 59: 2, 391.

<sup>60</sup> *ibid.* 392, quoting *Ashcroft v. Free Speech Coalition* 2002.

<sup>61</sup> Al-Alosi, 96.

<sup>62</sup> Al-Alosi, 96-97.

<sup>63</sup> *ibid.* 97.

<sup>64</sup> Levy, 319.

whether the causal relationships surveyed in our discussion of APF arguments actually hold. Moreover, pedophiles, when interviewed in studies, sometimes assert that viewing child abuse materials has a cathartic effect that “prevents them from acting upon their urges” or that they “have never consumed child abuse material before they molested a child.”<sup>65</sup>

Nevertheless, the phenomenon of viewing child sexual abuse materials is different from the phenomenon of viewing other forms of pornography because the former is “a valid indicator of paedophilia,” which is to say that the people who view these materials are pedophiles.<sup>66</sup> And while many pedophiles do not commit contact crimes, we should be skeptical of the trustworthiness of answers to self-report questionnaires where pedophiles are asked to admit to committing additional crimes—especially since “studies reveal that convicted child abuse material offenders often have committed child sexual abuse that goes unreported to authorities.”<sup>67</sup> The best evidence we have suggests that pedophiles who assert that viewing child sexual abuse materials have a cathartic effect that decreases the likelihood that they will sexually assault actual children are merely offering rationalizations of bad behavior. This is because “repeated exposure to child abuse material may cause desensitization” and normalization: behaviors that once seemed shocking and evoked a strong emotional response subsequently are seen as normal after habituation.<sup>68</sup>

As a result of desensitization and normalization, pedophiles reach a point at which “the images are no longer sufficient to meet the viewers’ sexual needs,” leading them to seek out more extreme forms of pornography and eventually an act of sexual assault against a real child.<sup>69</sup> Given the efficiency with which CAIP can be generated and the extent to which machine learning can expedite the process of creating desired content, it is plausible that pedophiles will progress through these stages of desensitization and normalization much more quickly than they already do. Thus, CAIP may, again, amplify a feature that constitutes one, of many, reasons why pornographic depictions of actual child sexual abuse are wrong.

Al-Alosi’s judgment is that “the research concerned with the impact of viewing child abuse material indicates the [fictional child pornography] can be harmful when consumed by sex offenders and paedophiles.”<sup>70</sup> Note, however, that her project concerned comic books and cartoons: fantasy materials that were not indistinguishable from pornographic depictions of actual events. As forms of CAIP, Dennis’s depictions in C and D may actually be indistinguishable from pornographic depictions of actual events, which would mean that they are perceptually indistinguishable from pornographic depictions of actual child sexual assault. Since the latter are even more strongly tied to sexual harms to actual children than non-indistinguishable fantasy materials Al-Alosi considered, our case is even stronger than hers.

The challenge we face, thus, is that there are good reasons to ban the conduct described in C and D based on a plausible interpretation of the harm principle, but such a ban would amount to a restriction on the simple possession of a self-created image that depicts the contents of one’s own

---

<sup>65</sup> Al-Alosi, 99-100.

<sup>66</sup> *ibid.* 100.

<sup>67</sup> *ibid.* 99.

<sup>68</sup> *ibid.* 100.

<sup>69</sup> *ibid.* 100.

<sup>70</sup> *ibid.* 106.

fantasies within the privacy of one's own home. Up until this year, an intervention into such a private domain would have seemed entirely beyond the scope of state intervention in a country like the United States, and this is because any reasonable interpretation of a right to freedom of expression should surely guarantee a right to come up with an original idea, create a depiction of that idea, and then destroy it without ever showing it to anyone else or using it for commercial gain. But due to the perceptual indistinguishability, automated and highly efficient production methods, unbounded nature, and customizability of CAIP, we have now reached a point where the exercise of such a right will plausibly result in tangible harms to children.

## **V. Conclusion**

Upon first glance, CAIP may not seem to be a novel nor ethically interesting phenomenon. In fact, CAIP might seem *prima facie* less ethically problematic than traditional pornography because the former can be created without exposing performers to physical harms. One might also think that deepfakes are the only instances of CAIP that raise serious ethical challenges, but that even these are no more novel than a peeping tom who illegally films an unsuspecting victim or an ex-partner who discloses sexually explicit images shared privately during happier times. It should be apparent by now that this deflationary attitude is false. Due to its five features, CAIP is fundamentally different from any pornography that has come before it, and the ways in which it is different have moral and legal implications that deserve our attention.

We have argued that CAIP is likely to be more addictive than traditional pornography because it can be efficiently generated and modified by machine learning algorithms that evaluate, at an unprecedentedly fine-grained level, which specific features cause specific people to persist in their consumption of it. This new form of addiction threatens to undermine people's agency, and it offers corporate interests a powerful tool to exploit our most powerful desires and psychological vulnerabilities. We have also argued that feminist arguments against pornography are amplified through CAIP because it can create novel forms of objectification, vastly expand the kinds and degrees of physical harm that can be depicted photorealistically, allow men to sexualize depictions of the very same women with whom they will interact in their lives, and lead to a world where some men spend far more time interacting with such depictions of particular colleagues, classmates, neighbors, etc., than they spend actually interacting with such real people.

We have also articulated a number of novel legal implications of CAIP. Existing and proposed laws governing deepfakes run the risk of either not regulating unmatched deepfakes or restricting speech that ought to be protected. And pure non-indexical CAIP is even more challenging to regulate than that. On the one hand, there are good reasons to believe that the simple possession of such CAIP can push pedophiles to sexually abuse actual children. On the other hand, it is implausible that banning the simple possession of CAIP, without an intent to distribute, is consistent with past decisions of the United States Supreme Court.

In this way, we have articulated an account of an undertheorized phenomenon and argued that it deserves attention. It is our hope that philosophers, legal theorists, psychologists, and other researchers will accept our invitation to take up projects that constitute next steps in this inquiry.