

**RAPID
ASSESSMENT
REPORT**

THE HIGH STAKES OF AI ETHICS:
**EVALUATING OPENAI'S
POTENTIAL SHIFT TO "NSFW"
CONTENT AND OTHER CONCERNS**

This report addresses lack of dataset transparency and the very likely unethical inclusion of sexually explicit material in OpenAI's pre-training and training datasets, risks associated with OpenAI's partnership with Reddit, and threats of harm posed by OpenAI's proposed "NSFW" rule change.

July 31, 2024

Prepared By

Lisa Thompson, Vice President
Research Institute
National Center on Sexual Exploitation

Presented By

Kindsey Chadwick, Interim President
Dawn Hawkins, CEO



NATIONAL CENTER ON
SEXUAL EXPLOITATION

The High Stakes of AI Ethics: Evaluating OpenAI’s Potential Shift to “NSFW” Content and Other Concerns

National Center on Sexual Exploitation

The National Center on Sexual Exploitation (NCOSE) is the leading organization preventing sexual abuse and exploitation at mass scale by eliminating institutional practices and societal norms that perpetuate these harms. As leading experts in combating sexual exploitation, NCOSE is uniquely positioned to assist in the critical endeavor of building AI that helps the world. Our decades of experience, combined with extensive research, direct engagement with hundreds of survivors, and advocacy with many other tech giants equips us with a profound understanding of the impacts—both positive and negative—of technology on society.

Recommended Citation:

Lisa Thompson, *The High Stakes of AI Ethics: Evaluating OpenAI’s Potential Shift to “NSFW” Content and Other Concerns* (Washington, DC: National Center on Sexual Exploitation, July 24, 2024).

Table of Contents

Why This Rapid Assessment?	4
OpenAI’s Mission and Commitments	4
OpenAI’s May 8, 2024, “Model Spec”	4
The Promise and the Peril of AI	4

PART I: LACK OF TRANSPARENCY & UNETHICAL INCLUSION OF SEXUALLY EXPLICIT MATERIAL IN DATASETS

1. OpenAI’s Lack of Dataset Transparency, Concerns Regarding CLIP & DALL-E, and Unethical Dataset Content	6
A. Rape Is a Crime, <i>Not</i> “Training Material”	8
B. The Impact of Hardcore Pornography and CSAM on Datasets	8
2. The Massive Problem of Nonconsensually Distributed Sexually Explicit Material of Adults on the Surface Web and Its Implications for Pre-training and Training Datasets	10
A. Nonconsensually Distributed Sexually Explicit Material Exists Widely on the Surface Web.....	10
B. Some Pornography Features Children and/or Adults Victimized by Sex Trafficking— Even on “Mainstream” Pornography Sites	12
C. Commercially Produced Material is Protected by Copyright	12
D. Lack of Age Verification.....	13

PART II: THE PERILS OF PARTNERSHIP WITH REDDIT

1. Reddit’s Dirty Dozen List Membership.....	14
2. Reddit’s Data Riddled with Sexual Exploitation Material	14

PART III: “NSFW” PROPOSED RULE CHANGE

1. Failure to Define “NSFW”	15
2. OpenAI’s Current Usage Policy and “NSFW” Rule Change Proposal Reflect a Naïve View of Corporate Actors and User Capacity for Abuse of AGI	15
A. Problems with OpenAI’s Current Usage Policy	16
B. The Current Impact of AI on Sexual Exploitation	18
3. How “NSFW” Material (Hardcore Pornography of Adults) Creates Widescale Harm	21
A. Child Sexual Abuse and Child Sexual Abuse Material.....	21

B. Sexual Violence.....	24
C. Sexually Transmitted Infections	26
D. Mental Health.....	27
E. Addiction-related Brain Changes	28
CONCLUSION	30
APPENDIX A.....	32
ENDNOTES.....	40

Why This Rapid Assessment?

OpenAI released model specifications on May 8, 2024, stating that it is exploring whether it can “responsibly provide the ability to generate NSFW content in age-appropriate contexts through the API and ChatGPT.” In response to this alarming development, as well as daily witnessing AI’s role in exacerbating sexual exploitation, NCOSE’s Research Institute has prepared this rapid assessment, *The High Stakes of AI Ethics: Evaluating OpenAI’s Potential Shift to “NSFW” Content and Other Concerns*, to inform OpenAI, other tech sector leaders, technology ethicists, and others about the perils of this reckless shift of direction and to raise other concerns.

OpenAI’s Mission and Commitments

OpenAI’s stated mission is “to ensure that artificial general intelligence benefits all of humanity.”¹ The company has also stated that it works to “make artificial general intelligence (AGI) aligned with human values and human intent,”² and formed a “Superalignment” team,ⁱ whose purpose was to develop “scientific and technical breakthroughs to steer and control AI systems much smarter than us.”³ OpenAI participated in the AI Seoul Summit, where it pledged to publish safety frameworks setting out how it will measure the risk of its models, and disable them when risks are severe and “intolerable.”⁴ Additionally, OpenAI publicly committed to principles outlined in *Safety by Design for Generative AI: Preventing Child Sexual Abuse* created by the advocacy organizations Thorn and All Tech is Human.⁵

OpenAI’s May 8, 2024, “Model Spec”

Despite its mission and commitments, on May 8, 2024, OpenAI released a “Model Spec” document in which OpenAI shockingly signaled that it is exploring changes to its current “Don’t respond with NSFW content” rule.⁶ Specifically, they stated: “We believe developers and users should have the flexibility to use our services as they see fit, so long as they comply with our usage policies. *We’re exploring whether we can responsibly provide the ability to generate NSFW content in age-appropriate contexts through the API and ChatGPT.*” (emphasis added)

The Promise and the Peril of AI

The National Center on Sexual Exploitation recognizes the incredible promise of AGI and the contributions of OpenAI to this rapidly developing field. OpenAI’s technology will help solve

ⁱ As we were in the process of writing this rapid assessment, both leaders of this team, Ilya Sutskever and Jan Leike, resigned from OpenAI and OpenAI disbanded its Superalignment team.

some of the most pressing issues the world faces today and could help solve problems related to sexual exploitation—if it so chooses.

However, it is impossible to ignore the **catastrophic impact** caused by the misuse of AI tools in the realm of sexual exploitation. We cannot sugarcoat the reality: the misuse of AI tools has unleashed massive sexual harm, impacting thousands who will suffer the consequences for their entire lives. **OpenAI’s proposed “NSFW” rule change, partnership with Reddit, and its lack of transparency regarding its pre-training and training datasets are poised to further these harms.**

Yet, time remains for OpenAI to change course and prioritize safety. OpenAI itself has expressed a desire for AGI to benefit humanity and “to understand user and societal expectations for AGI models.”⁷ Accordingly, the National Center on Sexual Exploitation (NCOSE) submits this rapid assessment to OpenAI in the hope that it, as well as other AI industry leaders, will take immediate and tireless action to mitigate the risks of adding further to the wave of AI-related sexual exploitation even now unfolding through the misuse of AI.

The NCOSE Research Institute’s rapid assessment outlines three areas of major concern with OpenAI’s practices:

- 1) lack of dataset transparency and the very likely unethical inclusion of sexually explicit material in its pre-training and training datasets**
- 2) perils of OpenAI’s partnership with Reddit due to Reddit’s data being riddled with sexually explicit, as well as sexual exploitation material**
- 3) threats of harm posed by OpenAI’s potential “NSFW” rule change pertaining to the generation of sexually explicit material/hardcore pornography.**

PART I:

LACK OF TRANSPARENCY & UNETHICAL INCLUSION OF SEXUALLY EXPLICIT MATERIAL IN DATASETS

1. OpenAI's Lack of Training Dataset Transparency, Concerns Regarding CLIP & DALL-E, and Unethical Dataset Content

The sources of OpenAI's training datasets are undisclosed.⁸ This ambiguity, coupled with the sheer volume of images required for creating foundation models, makes it highly likely that images already within its pre-training and training datasets contain nonconsensual and illegal material.

Illustrating this probability, Prabhu and Birhane⁹ found verifiable voyeuristic, pornography of individuals, mostly women, in their analysis of images contained in ImageNet,ⁱⁱ which is purportedly “one of the most widely used datasets in machine learning.”¹⁰ Even more disconcerting, they were able to “easily map the victims, most of whom are women, in the pictures to ‘real-world’ identities of people belonging to a myriad of backgrounds including teachers, medical professionals, and academic professors using reverse image search engines. . . .”¹¹ Thus, ImageNet violated the consent and privacy of these individuals, and thereby utilized unethical content—content that violates basic principles for the protection of human subjects in research.ⁱⁱⁱ **Models that train on ImageNet are abuse-trained.**

Given ImageNet's prominent place in the historical development of AGI and the rush to release AGI models by developers despite the lack of strong governing principles

ⁱⁱ We use ImageNet as an illustration of an early and very influential dataset that was corrupted by unethical material. The extent to which OpenAI has used or borrowed from ImageNet in the development of its AGI is not publicly known.

ⁱⁱⁱ For an articulation of these principles see: The National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, *The Belmont Report. Ethical Principles and Guidelines for the Protection of Human Subjects of Research*, (Washington, DC: U.S. Department of Health, Education, and Welfare, 1978), https://www.hhs.gov/ohrp/sites/default/files/the-belmont-report-508c_FINAL.pdf (accessed May 15, 2024).

regarding its development, NCOSE has strong reason to be concerned about the material in the datasets used to pre-train and train OpenAI's models.

OpenAI reports that its DALL-E 2 was trained using a dataset of hundreds of millions of text-image pairs filtered to exclude violent and sexual images,¹² or what OpenAI has called "the most explicit content."¹³ While **we fully endorse the exclusion of violent and sexually explicit images from DALL-E (and other models) training datasets**, NCOSE has multiple concerns. For instance, there is a lack of transparency about how OpenAI defines "the most explicit content." Does "the most explicit content" encompass sexualized but not "explicit" images of children, hardcore pornography, violent sex acts (e.g., hair pulling, slapping, punching, strangulation, electric shocking, torture), racism themed-pornography, etcetera? Further, who or what at OpenAI makes the determination of what constitutes "the most explicit content"?

Of further concern, in its evaluation of DALL-E 2, Common Sense Media reported that attempts to filter sexually explicit material from DALL-E's pre-training data resulted in a reduction in the frequency of the keyword "woman" by 14%, whereas the explicit content filter only reduced the frequency of the keyword "man" by 6%. As they observed, "In other words, OpenAI's attempts to remove explicit material removed enough content representing women that the resulting data set was significantly overrepresented by men. This offers perspective on how many images on the internet contain sexual content of women."¹⁴ Indeed, it does. Please see the discussion beginning on page 10 for more on the profusion of pornography on the Internet.

Moreover, problems regarding the representation of women by AI models derive not only from the potential exclusion of women from datasets, but from the fact that sexually explicit data remaining in datasets alters the way in which AI learns to "see" women and thus portray them in text-to-image generative-AI. As Common Sense Media's review of DALL-E further noted:

DALL-E continues to demonstrate a tendency toward objectification and sexualization. This is especially the case with inappropriate sexualized representations of women and girls, even with prompts seeking images of women professionals. This perpetuates harmful stereotypes, unfair bias, unrealistic ideals of women's beauty and "sexiness," and incorrect beliefs around intimacy for humans of all genders.¹⁵

Birhane, Prabhu, and Kahembwe elaborate on the problems with the content and curation of datasets on which machine learning models are trained, saying:

The rise of these gargantuan datasets has given rise to formidable bodies of critical work that has called for caution while generating these large datasets. These address concerns surrounding the dubious curation practices used to generate these datasets, the sordid quality of alt-text data available on the world wide web, the problematic content of the CommonCrawl dataset often used as a source for training large language models, and the entrenched biases in large-

scale visio-linguistic models (such as *OpenAI's CLIP model*) trained on opaque datasets (WebImageText). In the backdrop of these specific calls of caution, we examine the recently released LAION-400M dataset, which is a *CLIP-filtered dataset* of Image-Alt-text pairs parsed from the Common-Crawl dataset. ***We found that the dataset contains, troublesome and explicit images and text pairs of rape, pornography, malign stereotypes, racist and ethnic slurs, and other extremely problematic content.***¹⁶ (bold emphasis added)

It is extremely disconcerting that OpenAI's CLIP neural network, used to help create LAION-400M dataset, resulted in the inclusion of metadata of rape, pornography, and other "extremely problematic content" within LAION-400M. How could this be unless CLIP itself fails to adequately filter such material? Because of the apparent failure of OpenAI's CLIP to adequately filter LAION-400M (as well as LAION-5B to be discussed shortly), and DALL-E's poor performance with respect to its generation of images depicting women, we cannot merely "trust" that OpenAI training data is free of sexually abusive and exploitative material. OpenAI must provide *evidence* that its data training sets are free of sexual abuse and exploitation material.

A. Rape Is a Crime, Not "Training Material"

Since rape material was identified in the CLIP-filtered LAION-400M data set, we wish to point out that rape is "penetration, no matter how slight, of the vagina or anus with any body part or object, or oral penetration by a sex organ of another person, without the consent of the victim" and is illegal.¹⁷ Images or recordings of rape, incidents of severe physical, psychological, and sexual trauma, forever memorialize moments of terrifying sexual violence, and their distribution online amplifies this violence by rendering someone's experience of sexual violation into masturbatory material for a global audience. Inclusion of such material (or its metadata) in any OpenAI datasets and/or models, or failure by OpenAI tools to filter out all such material, violates the most basic precepts of human rights and dignity. Any inclusion of images or videos depicting rape, an innately violative sexual act, in pre-training or training datasets is inherently unethical.

The potential use of AGI to generate material depicting rape or sexual violence is, likewise, unconscionable. AI-generated material depicting rape, even though it uses representations of human beings, not depictions of actual humans, nevertheless perpetuates harmful societal norms and attitudes toward sexual violence.

B. The Impact of Hardcore Pornography and CSAM on Datasets

Likewise, large-scale datasets, such as those derived from the Common Crawl, have been shown to include sexually explicit content.¹⁸ Luccioni and Viviano

note that even a small percentage of such content can have major consequences, explaining:

... the type of language and content contained on adult websites can have harmful repercussions. For instance, the prevalence of sexual violence towards women, especially towards women of color, on adult websites (Foubert et al., 2019; Shim et al., 2015; Fritz et al., 2020) **may contribute to further dissemination and amplification of these biases in downstream models.** As modern language models have no way to evaluate generation appropriateness, **models trained with even a small proportion of these undesirable inputs cannot be guaranteed to avoid generating outputs with similar biases if presented with a specific context or prompt.** This is a risk that is important to mitigate in applications, where the general-purpose language models can end up being used in applications used by sensitive groups in professional contexts or minors, such as chatbots and toys.¹⁹ (emphasis added)

We note, too, that LAION-5B, which supplanted LAION-400M, was also CLIP-filtered.²⁰ An analysis by Daniel Thiel of the Stanford Internet Observatory confirmed that, nevertheless, **LAION-5B was trained on “a wide array of content, both explicit and otherwise,”**²¹ and that “CSAM was identifiably present in the dataset.”²² The report also contained this alarming statement:

We find that having possession of a LAION-5B dataset populated even in late 2023 implies the possession of thousands of illegal images—**not including all of the intimate imagery published and gathered non-consensually,** the legality of which is more variable by jurisdiction.²³ (emphasis added)

Given the serious ethical and legal deficiencies, especially as pertains to matters of child sexual abuse material and nonconsensually distributed sexually explicit material of adults (**also known as image-based sexual abuse (IBSA)**)^{iv}, manifested with ImageNet, CLIP, LAION-400M, LAION-5B, and the Common Crawl, greater transparency into the datasets created and/or used by AGI creators, and how they are curated, is imperative. While it is encouraging that a news report indicates that DALL-E does not train on any LAION dataset,²⁴ **OpenAI’s CLIP was used to filter both the LAION 400-M and 5B releases,**

^{iv} The National Center on Sexual Exploitation defines image-based sexual abuse as the sexual violation of a person committed through the abuse, exploitation, or weaponization of any image depicting the person. An “image” is any visual depiction or representation of a person—including but not limited to materials such as photographs, videos, edited/altered images, or personal representations in virtual reality or online gaming. IBSA encompasses a wide range of abusive activities including the creation, theft, extortion, threatened or actual distribution, or any use of images for sexual purposes without the meaningful consent of the person/s depicted or for purposes of sexual exploitation. It also includes sexual violence or harassment committed towards a person’s representation (e.g., a person’s avatar) in virtual reality or online gaming.

thus OpenAI is in part responsible for the LAION datasets, which are laden with extremely problematic—including illegal material—and which are in wide use by the AI-development community.

The consequences of this failure within the AGI development community are devastating, and as has been noted elsewhere, **“We are now faced with the [mammoth] task of trying to mitigate the material present in extant training sets and tainted models. . . .”**²⁵

OpenAI must take accountability for its role in this failure and develop robust measures to ensure that mistakes of such import and magnitude are not repeated.

2. The Massive Problem of Nonconsensually Distributed Sexually Explicit Material of Adults on the Surface Web and Its Implications for Pre-training and Training Datasets

There is no means by which image dataset creators can authenticate the *meaningful consent* of adults depicted in sexually explicit material presently on the Internet on a scale that matches the datasets themselves. “Meaningful consent” means the documented, informed consent of all persons depicted in sexually explicit content, for each discreet image or recording, before the creation and dissemination of such content, and the right of persons depicted to withdraw consent and remove sexually explicit content at any time thereafter. Any use of a dataset that incorporates sexually explicit material without meaningful consent is unethical.

A. Nonconsensually Distributed Sexually Explicit Material Exists Widely on the Surface Web

The nonconsensual distribution of sexually explicit material, otherwise known as image-based sexual abuse, is an entrenched problem on the Internet, with such content being widely and readily available on the surface web and intermixed with commercially produced pornography.²⁶ This problem was largely created by pornography tube sites like Pornhub and XVideos, which, for more than a decade, allowed anonymous user-generated content to be uploaded to and downloaded from their websites.²⁷ Some of this user-generated material included forms of IBSA such as recorded sexual violence (e.g., recordings of sexual assault, sexual assaults of women while drugged or intoxicated, the rape of sex trafficking

victims), video voyeurism,^v and sexually explicit material that was nonconsensually uploaded by former romantic partners, third-parties, or which was obtained via hacking. Illustrating the perniciousness of the problem, in December 2020, Pornhub removed 10.6 million videos uploaded by unverified users, or roughly 80% of its content,^{vi} after an expose in *The New York Times* revealed that the site was riddled with CSAM and IBSA material.²⁸ **The mass trafficking of image-based sexual abuse material enabled by the unverified user uploads and downloads on “mainstream” pornography platforms with billions of monthly users resulted in IBAs proliferation on the surface web.**

The harms to those who have experienced this form of victimization are wide ranging and enduring. In a report following an investigation into Pornhub and its parent company MindGeek (now Aylo) conducted by a committee of the Canadian Parliament, the lawmakers reported:

... survivors shared the trauma of having abusive images of them uploaded online without their knowledge or consent—many of which depicted activities undertaken or recorded without their consent, or both. Some explained how the upload of their images revictimized them; it allowed pornography sites to profit from their violation and allowed viewers to take pleasure in, comment on, download and re-upload that experience. In addition, most survivors spoke of having their identifying information posted online, being bullied, harassed, or followed in person by their abusers and other people who had seen them online. **Many explained that they had developed mental and physical illnesses that prevented them from functioning in their daily lives, and most had considered or attempted suicide.**²⁹

We raise this matter because crawls or scrapes of images from the web are very likely—if not guaranteed—to incorporate CSAM and IBSA material, material that mainstream pornography companies allow to metastasize on their platforms. Neither humans nor machines can, with complete accuracy, distinguish between the two—especially since some nonconsensually distributed sexually explicit material is initially created consensually. As noted above, such material has already been found in large-scale datasets used for AGI model training undermining the ethical integrity of the datasets and the AGI they are used to create.

^v The National Center on Sexual Exploitation defines video voyeurism as the surreptitious recording of images or videos of clothed people while engaged in public activities in order to obtain images of their private body parts (e.g., “down blousing,” and “upskirting”) or the recording of persons engaged in private activities (e.g., changing clothes, using the toilet or showering, having sex in private) without their knowledge.

^{vi} We do not assert that all 10.6 million videos were CSAM or IBSA, but report this figure to demonstrate the overwhelming presence of unverified material on Pornhub’s platform at that time.

The incorporation of CSAM and/or IBSA into AGI training datasets not only further exploits those victimized by it but also turns their sexual abuse and exploitation into “neural learning” for machines. This is unconscionable.

B. Some Pornography Features Children and/or Adults Victimized by Sex Trafficking—Even on “Mainstream” Pornography Sites

“Women, men, girls, and boys are sex trafficked for the production of pornography.”³⁰ Illustrating this point, the operators of Girls Do Porn and Czech Casting, two “*verified*” Pornhub content partners, were arrested or convicted on charges of sex trafficking,³¹ and in 2023 Aylo Holdings (Pornhub’s parent company) admitted to federal prosecutors that it profited from sex trafficking.³² **This further illustrates the futility of attempting to classify sexually explicit material into that which is acceptable for AGI training and that which is not.** Without knowing the person depicted and possessing information about their personal history and current circumstances, one cannot merely look at pornography and know whether or not the individual pictured has been exploited by sex traffickers. This holds true for pornography depicting adults writ large.

Further we note with horror that, “Images of sex trafficked women have been harvested from pornography sites and used to create so-called ‘artificial’ pornography.”³³ Journalists discovered that both Girls Do Porn and Czech Casting material is in a dataset being used to create AI-generated (i.e., “deepfake”) pornography.³⁴ We presume that this dataset has not been used by OpenAI, but we point out this case to demonstrate the highly sexually exploitative purposes for which AI is already being employed. It also further evidences the highly exploitative nature of pornography production and the impossibility of discerning which pornography was made with exploitation and which was not.^{vii}

C. Commercially Produced Material is Protected by Copyright

In addition to the range of concerns about violations of copyright that overshadow the development of AGI, OpenAI should be aware that commercially produced pornography is protected by the Digital Millennium Copyright Act (DMCA) of 1998. NCOSE also alerts OpenAI to the fact that pornography tube sites, from their very inception in 2007, have depended on large amounts of pirated pornography. One journalist put it this way, “Porn doesn’t suffer from a piracy problem . . . it actively relies upon it.”³⁵ The issue continues to plague those creating pornography today. For instance, subscribers of OnlyFans are notorious for

^{vii} It is important to note that even those who are not engaged in pornography production due to means of force, fraud, and coercion, are still subject to sexual abuse and exploitation. For instance, see Meghan Donevan, “‘In This Industry, You’re No Longer Human’: An Exploratory Study of Women’s Experiences in Pornography Production in Sweden,” *Dignity: A Journal of Analysis of Exploitation and Violence* 6, no 3 (2021): <https://digitalcommons.uri.edu/cgi/viewcontent.cgi?article=1271&context=dignity>.

ripping the content pornography performers monetize there and distributing it on other platforms. A Motherboard investigation into the issue “uncovered an entire supply chain of people stealing” performers’ content “using scraping programs, without permission, in some cases by the hundreds of terabytes, and distributing it on other adult sites or selling scraping services through Discord.”³⁶ **Thus, the odds are exceedingly likely that at least some of the hardcore pornography currently within some data training sets is pirated material.**

Nonauthorized use of commercially produced pornography not only violates the DMCA, but also violates the rights of pornography performers to meaningful, informed consent about how their material will be used. **Agreeing to perform sex acts via information communications technology for compensation is not the same as agreeing to perform sex acts to train the neural networks that form AGI models—models from which OpenAI and its shareholders will exponentially profit.** The use of commercially produced content without meaningful informed consent in any pre-training or training datasets for AGI is a blatant violation of the rights of pornography performers.

D. Lack of Age Verification

The technological capacity to age-gate minors’ access to age-restricted material utilizing facial age *estimation* has improved.³⁷ However, machines will not be able to ethically *verify* the age of persons in sexually explicit images, because to do so would require the *positive* identification of *every* person in *every* image, that is, *facial* recognition. The point is that the utilization of sexually explicit material of adults obtained via large Internet crawls or scrapes for use in pre-training or training datasets cannot be ethically undertaken. **To do so requires the meaningful consent of those depicted, yet meaningful consent cannot be obtained for the process of positively identifying those depicted, as it cannot be obtained in advance of that process being undertaken.** (In contrast, hashing has enabled the identification of known CSAM material without the risk of facial recognition and should also be utilized to also classify known AI-CSAM,^{viii} IBSA, and AI-IBSA.^{ix})

In light of the above concerns, it is patently obvious that the *ethical* inclusion of sexually explicit material depicting adults obtained via web crawls or scrapes in the pre-training or training datasets for AGI is an impossibility.

^{viii} The National Center on Sexual Exploitation utilizes the term AI-CSAM to refer to images or videos depicting the sexual abuse of children created by any generative-AI or AI-related technologies such as machine learning, deep learning, neural networks, computer vision, etc.

^{ix} Likewise, the National Center on Sexual Exploitation uses the term AI-IBSA to refer to images or videos depicting any form of image-based sexual abuse created by any generative-AI or AI-related technologies such as machine learning, deep learning, neural networks, etc.

PART II: THE PERILS OF PARTNERSHIP WITH REDDIT

1. Reddit’s Dirty Dozen List Membership

We are *deeply* concerned by OpenAI’s recent partnership with Reddit.³⁸ **Reddit is a member of The Dirty Dozen List 2024, a list of major contributors to sexual exploitation published annually by the National Center on Sexual Exploitation.**³⁹ As reported earlier this year, Reddit “is a hub of image-based sexual abuse, hardcore pornography, prostitution—which very likely includes sex trafficking—and overt cases of child sexual exploitation and child sexual abuse material.”⁴⁰ Reddit has been a member of The Dirty Dozen List every year since 2021.

2. Reddit’s Data Riddled with Sexual Exploitation Material

The appeal of Reddit’s data as a source of data for ChatGPT is understandably tantalizing, given its 1 billion cumulative posts and 73 million average daily unique visitors.⁴¹ **However that data is riddled with sexually explicit and exploitative content.** NCOSE has previously notified Reddit of this concern, writing to their executives with alarm regarding their partnership with Google to train its AI models, explaining, “. . . without stringent safeguards to filter out image-based sexual abuse and CSAM, the potential for Google’s AI models to be trained on the pervasive amount of sexually exploitative and potentially illegal material on Reddit is much too high.”⁴² **OpenAI’s partnership with Reddit to train models like ChatGPT is at high risk of replicating errors like those of Stable Diffusion’s LAION-5b image-text model.**

The evidence of sexually explicit, as well as sexual exploitation material on Reddit’s platform has been well documented elsewhere. Please see Appendix A to view NCOSE’s 2024 letter to Reddit for a full elucidation of those concerns.

AI training on Reddit data without very robust filtering will result in an abuse-trained model.

PART III:

“NSFW” PROPOSED RULE CHANGE

1. Failure to Define “NSFW”

As currently formulated, OpenAI’s rule “Don’t respond with NSFW content” uses the acronym “NSFW” for “not safe for work”—a slang term we assume you use to refer to sexually explicit material depicting adults,^x as well as other subject matter unsuitable for professional and other settings (e.g., racial slurs, gore, profanity, sexual violence, etc.). The use of slang terminology like NSFW to refer to the serious subject of what kind of material OpenAI will empower its users to create belittles the gravity of the issues involved. Hardcore pornography (obscenity), as well as subjects like racism, extreme violence, and sexual violence, are not trivial matters but are social issues that deeply impact the health and well-being of our world. Such vagary also creates confusion for OpenAI users. What precisely is meant by “NSFW” is open to debate as OpenAI’s Universal Policies⁴³ provide no explanation.

Therefore, NCOSE requests that OpenAI invest considerable time and thought in defining types of currently violative “NSFW” content and making those definitions publicly available so that users can better understand the parameters of appropriate use of OpenAI tools.

2. OpenAI’s Current Usage Policy and “NSFW” Rule Change Proposal Reflect a Naïve View of Corporate Actors and User Capacity for Abuse of AGI

OpenAI’s proposed rule change in its May 8 “Model Spec” document pertaining to “NSFW” material states, “*We believe developers and users should have the flexibility to use our services as they see fit, so long as they comply with our usage policies.*” (emphasis added)

Such an attitude is naïve at best and an open invitation to abuse at worst.

^x 18 U.S. Code § 2256 defines “sexually explicit conduct” as “actual or simulated (i) sexual intercourse, including genital-genital, anal-genital, or oral-anal, whether between persons of the same or opposite sex; (ii) bestiality; (iii) masturbation; (iv) sadistic or masochistic abuse; or (v) lascivious exhibition of the anus, genitals, or pubic area of any person. See <https://www.law.cornell.edu/uscode/text/18/2256>. In this document we use the terms “sexually explicit” and “hardcore pornography” interchangeably.

To begin, the tech industry’s track record regarding monitoring and enforcement of its usage policies has categorically and unquestionably demonstrated that neither they nor their users respect “terms of use.” Perhaps OpenAI is unaware of the January 31, 2024, Senate Judiciary Hearing at which executives of Discord, Meta, Snapchat, TikTok, and X were “called on the carpet” for their companies’ abject failures to enforce their own policies and to keep their users safe from those who utilize their systems for purposes including sexual harassment and criminal forms of sexual exploitation.⁴⁴ Time and time again, NCOSE and others have documented that technology companies like Apple,⁴⁵ Cloudflare,⁴⁶ Discord,⁴⁷ Meta,⁴⁸ Microsoft’s GitHub,⁴⁹ Reddit,⁵⁰ X,⁵¹ and others have shown that they are unwilling to enforce their own policies and to prioritize the safety and wellbeing of their users.

Considering the abysmal track record of OpenAI’s industry peers, NCOSE has little faith that OpenAI’s commitment to enforcing its “usage policies” is greater than its commitment to industry share and financial gain. NCOSE will be overjoyed if OpenAI proves us wrong. To do so, addressing gaps and the lack of clarity and forcefulness in its current Usage Policy must be an OpenAI priority.

A. Problems with OpenAI’s Current Usage Policy

At present, Open AI’s Usage Policy lacks forcefulness and clarity in multiple respects.

- I. We have already pointed out its lack of specificity regarding the definition of so-called “NSFW” material.
- II. OpenAI’s “Universal Policies” currently state, “*To maximize innovation and creativity, we believe you should have the flexibility to use our services as you see fit, so long as you comply with the law and don’t harm yourself or others*”⁵² (emphasis added). From the outset OpenAI signals that it places “innovation and creativity” on par with safety when safety must take precedence.

Its own former and current employees share a similar concern.

Jan Leike, former Head of Alignment at OpenAI, stated:

- “. . . over the past years, safety culture and processes have taken a backseat to shiny products.”⁵³
- “Over the past few months my team has been sailing against the wind.”⁵⁴
- “I joined because I thought OpenAI would be the best place in the world to do this research. However, I have been disagreeing with

OpenAI leadership about the company's core priorities for quite some time, until we finally reached a breaking point."⁵⁵

Daniel Kokotajlo, former researcher, governance division, OpenAI, commented:

- "OpenAI is really excited about building A.G.I., and they are recklessly racing to be the first there."⁵⁶
- The world isn't ready, and we aren't ready. . . . And I'm concerned we are rushing forward regardless and rationalizing our actions."

William Saunders, a former research engineer at OpenAI, said, "When I signed up for OpenAI, I did not sign up for this attitude of 'Let's put things out into the world and see what happens and fix them afterward.'"⁵⁷

Further, 11 former and current OpenAI employees have signed an open letter to "frontier AI companies" such as OpenAI, decrying the risks that AI systems pose to humanity.⁵⁸ That 11 former and current members of **OpenAI felt it necessary to publicly express their concern about its deficient safety practices is the billowing smoke that signals a flaming failure at OpenAI to prioritize safety.**

- III. OpenAI's guidance to "Comply with applicable laws"⁵⁹ falls short. If it is legal to create CSAM, AI-CSAM, IBSA, or AI-IBSA in a particular country, what then? For instance, creators of AI-CSAM are using a social media platform in Japan precisely because sexualized cartoons and drawings of children are not illegal there.⁶⁰
- IV. OpenAI's guidance to "Don't use our service to harm yourself or others" provides only examples involving physical harm to others, such as encouraging self-harm, developing or use of weapons, or injuring others. **Notably omitted from this list is *sexual violence*.** Additionally, harm to others, especially where the technology-related abuses of IBSA and AI-IBSA are involved, frequently includes psychological harms such as depression and anxiety that result from damage to one's relationships and reputation, as well as relentless public shaming and sexual harassment from the misappropriation of one's likeness. **The language here should be broadened to incorporate these concerns.**
- V. Item 3 of OpenAI's Usage Policy advises: "**Don't repurpose or distribute output from our services to harm others** – for example, don't share output from our services to defraud, scam, spam, mislead, bully, harass, defame, discriminate based on protected attributes, sexualize children, or promote violence, hatred or the suffering of others."⁶¹ This section of its policy is stronger than most, **yet NCOSE notes that there is no explicit reference**

restricting the distribution of material depicting sexual violence broadly or the distribution of AI-generated-IBSA (AI-IBSA) in particular.

- VI. In the current iteration of “Building with the OpenAI API Platform” and “Building with Chat GPT” subsections of OpenAI’s Usage Policy, **neither expressly prohibits the generation of AI-sexual violence or AI-IBSA.** Such prohibitions should also be added to OpenAI’s “Universal Policies.”
- VII. NCOSE commends OpenAI’s prohibition on the *repurposing* or *distribution* of output that sexualizes children. However, this must be extended to a **universal policy that prohibits the generation of output that sexualizes children, not merely its repurposing or distribution.**
- VIII. NCOSE also commends OpenAI’s “Building with the OpenAI API Platform” and “Building with Chat GPT” policies, which prohibit the building of tools that may be inappropriate for children, such as sexually explicit or suggestive content. This provision alone should nullify OpenAI’s proposal to remove restrictions on so-called NSFW material.
- IX. Additionally, NCOSE observes that absent from “OpenAI’s Approach to Frontier Risk” is any mention of the sexual harm associated with AGI. OpenAI’s “Preparedness Framework” recognizes risks in areas such as “cybersecurity, persuasion, chemical and biological threats, and autonomy.”⁶² While this framework was created to address the range of risks that were to be considered at the AI Safety Summit, yet, **nowhere** is there any recognition of the ongoing sexual exploitation that has arisen from AI or of the urgent need to stop it. **This signals to us that OpenAI and other AI sector leaders are trapped in thought bubbles that exclude awareness of, sensitivity to, and concern for issues related to sexual abuse, violence, and exploitation. This situation must be rectified with haste.** All preparedness frameworks must address sexual abuse, violence, and exploitation.

B. The Current Impact of AI on Sexual Exploitation

Furthermore, irresponsible developers and other individuals have set scruples aside and released a wave of AI-empowered tools that are wreaking havoc in society and causing irreparable damage. Evidence abounds of the reckless and nefarious uses to which AI is already being employed.

- I. **Chatbots:** Uncensored and loosely moderated AI chatbots have disseminated misinformation, provided advice on how to commit suicide,⁶³ and given instructions to a researcher posing as a 13-year-old girl on how to have sex with a 31-year-old man. ⁶⁴ ChatGPT has been manipulated **by**

online communities on Reddit to respond to sexually explicit prompts and create what one journalist called “wildly explicit smut.”⁶⁵ In fact, **researchers have identified 78 “jailbreak” prompts circulated for the very purpose of circumventing ChatGPT’s safety and moderation features.**⁶⁶

However, as OpenAI must be aware, ChatGPT can produce alarming results with no outside manipulation. For instance, it invented a story about a law professor, claiming that he had sexually harassed someone, and “cited” a non-existent article from *The Washington Post* as evidence.⁶⁷ This, sadly, is not an isolated incident. ChatGPT has made up false and defamatory information on several other occasions.⁶⁸

Additionally, Australian eSafety Commissioner Julie Inman Grant has expressed concern that predators will create chatbots for the purpose of automating “child grooming at scale.”⁶⁹

- II. **Nudify apps:** Nudify apps that strip people—overwhelming women and girls—of their clothes have spawned like a plague. Research has revealed that 34 providers of nudifying technology received more than 24 million unique visitors to their websites during the month of September 2023.⁷⁰ Further, “the volume of referral link spam for these services has increased by more than 2,000% **on platforms including Reddit** and X since the beginning of 2023,” and in September of 2023, there were at least 1 million users of 52 Telegram groups used to create AI-IBSA.⁷¹ As journalist Nicolas Kristof has documented, **some of the videos being “created are graphic and sometimes sadistic, depicting women tied up as they are raped or urinated on, for example. One site offers categories including ‘rape’ (472 items), ‘crying’ (655) and ‘degradation’ (822).”**⁷² Kristof also found one deepfake website that displays the official portrait of a female member of Congress, as well as 28 fake sex videos of her.⁷³ Add to this the crisis being faced by school districts across the US and elsewhere, as boys discover and use AI-powered apps to transform clothed pictures of their female classmates into nude images.⁷⁴
- III. **AI-generated Sexualized Images of Children:** AI-generated images of sexualized children are flooding social media apps like TikTok and Instagram. The material has been described this way:

The girls in the photos on TikTok and Instagram look like they could be five or six years old. On the older end, not quite thirteen.

They’re pictured in lace and leather, bikinis and crop tops. They’re dressed suggestively as nurses, superheroes, ballerinas and French maids. Some wear bunny ears or devil horns; others, pigtails and oversized glasses. They’re black, white and Asian, blondes, redheads

and brunettes. They were all made with AI, and they've become magnets for the attention of a troubling audience on some of the biggest social media apps in the world—older men.⁷⁵

The proliferation of this kind of material further normalizes the sexualization of children. Forums where it is traded share instructions on how to generate similar material and serve as gateways to CSAM.⁷⁶

- IV. **AI-Generated CSAM:** Underground CSAM trading networks are disseminating AI-generated CSAM. The Internet Watch Foundation (IWF) reported in 2023 that **one forum posted more than 20,000 images of AI-CSAM in a one-month period and that “there’s jubilation that fantasies can be made to order” among perpetrators.**⁷⁷ The IWF further reported AI-generated CSAM using images of known, real victims, “de-aging” of celebrities, AI-generated CSAM of celebrity children, nudifying of clothed children, and commercialization of all such material.⁷⁸ In a July 2024 report, the IWF found that 90% of AI-CSAM images assessed by IWF analysts were “realistic enough to be assessed as pseudo-photographs of children. . . .”⁷⁹ In addition, the proportion of material that was classified as the most severe, that is Category A (images depicting penetrative sexual activity; images involving sexual activity with an animal; or sadism), increased by 10 percentage points. The IWF also reported that the first AI-CSAM videos are now circulating.

Pedophiles are also accessing this material through mainstream platforms such as Patreon and using it to create “life-like images of child sexual abuse, including the rape of babies and toddlers.”⁸⁰ The National Center for Missing and Exploited Children (NCMEC) has experienced roughly 450 reports of AI-generated CSAM *per month* in the first quarter of this year.⁸¹ We note with deep grief that an estimated 300+ million children have been affected by online child sexual exploitation in the last 12 months.⁸² **AI-generated CSAM throws gasoline on this already raging problem, both normalizing CSAM and making the identification of child victims exceedingly more difficult.**

These problems are nothing short of a hellscape—a hellscape of the AI sector’s making.

It is against this backdrop of mammoth and out-of-control sexual exploitation generated and enflamed by AI that OpenAI says it is considering permitting the so-called “ethical” generation of “NSFW” material by its users!

NCOSE, therefore, asks, is OpenAI unconcerned with the scope of damage that has already been unleashed on the world by the open, rushed, and unregulated release of AI?

Is OpenAI willfully blind to the raging and uncontained problems that generative AI has already unleashed?

Is it not beneath OpenAI and its noble aspirations of bettering humanity to succumb to the demands of the basest users of AI technology?

Is “NSFW” material the purpose to which OpenAI will devote the talents of its employees and the most powerful technology in the world?

3. How “NSFW” Material (Hardcore Pornography of Adults) Creates Widescale Harm

Evidence from peer-reviewed research demonstrates that consumption of “NSFW” material (i.e., mainstream, hardcore pornography depicting adults) is associated with an array of adverse impacts that exacerbate global public health concerns, including: **1) child sexual abuse and child sexual abuse material, 2) sexual violence, 3) sexually transmitted infections, 4) mental health harms, and 5) addiction-related brain changes.**

A. Child Sexual Abuse and Child Sexual Abuse Material

Child sexual abuse experts have explained:

Lifetime exposure to child sexual abuse (CSA) and other forms of sexual harm (e.g., sexual exposure, sexual harassment, and Internet sex talk) affects approximately 10% of a nationally representative sample of U.S. children aged 0–17 years, including 12% of girls and nearly 8% of boys. Such exposure **significantly increases the likelihood of subsequent sexual and nonsexual revictimization for boys and girls and subsequent sexual offending for boys.**

CSA is among 24 global risk factors identified by the World Health Organization that substantively affect the global burden of disease, contributing an estimated **0.6% to the global burden of disease, or 9 million years of healthy life lost.**⁸³

It is vital to add to this fact that the production of **CSAM, by its very nature, involves child sexual abuse and exploitation of real children.** Reports of CSAM have grown exponentially in recent years, with 3,000 reports in 1998 growing to more than 1 million in 2014 to 18.4 million in 2018—those reports included more

than 45 million images and videos^{xi} flagged as child sexual abuse.⁸⁴ *The New York Times* called it an “almost unfathomable” increase in criminal behavior.⁸⁵

Hardcore pornography depicting adults or sexually explicit “NSFW” material exacerbates these staggering problems via the process of desensitization and escalation.

Desensitization and Escalation from Hardcore Pornography to CSAM and Contact Offenses

- A study of general Internet users assessed if a person’s “age of onset” of “adult” pornography use was a risk factor for future use of **“deviant pornography” (i.e., child and/or animal pornography)**. A survey of 630 subjects found that **individuals who engaged in adult pornography use at a younger age were more likely to transition to deviant pornography use, demonstrating desensitization to adult pornography over time. Those using both adult and deviant pornography were twice as likely to report age of onset between 12 and 18 years of age than those who only used adult pornography. No respondents reported using only CSAM.** This is indicative of a Guttman-like progression^{xii} in which individuals who consume CSAM also consume other forms of both deviant and nondeviant pornography, with CSAM use being more likely to occur **after** “nondeviant” pornography use.⁸⁶
- Similarly, a study of a general population of Internet users, using a different sampling method, replicated the study above, found that those who reported the use of CSAM were more likely to consume both adult and animal pornography rather than solely consuming CSAM. Additionally, **41% of those who used adult and deviant pornography (i.e., child and/or animal pornography) reported first using pornography when under 16 years of age, compared to only 30% of those only using adult pornography.** The study’s findings indicated a progression of pornography use in which deviant pornography use is more likely to occur **after** the use of nondeviant adult pornography. The researchers explained, “Preliminary findings suggest that deviant pornography use, specifically child pornography, may follow a Guttman-like progression for some individuals in that **they become desensitized and progress from nondeviant forms of pornography to more extreme genres**”⁸⁷ (emphasis added). Several additional studies

^{xi} We do not assert that all 45 million images are of individual children, recognizing that multiple images of a single child are often produced, or that all 45 million images are of “new” material as older CSAM is collected and traded among pedophiles and other collectors. Nevertheless, some of this vast quantity of material doubtlessly represents previously undocumented crimes of sexual abuse of children.

^{xii} Guttman scales measure progressively higher levels of attributes. “For example, a series of interview questions might ask, “Do you smoke?” “Do you smoke more than 10 cigarettes a day?” “Do you smoke more than a pack a day?” and so on” (Robert F. DeVellis, *Scale Development: Theory and Applications, Second Edition*, (Thousand Oaks: CA, Sage Publications, 2003), p. 73.

provide evidence that mainstream pornography serves as a gateway to CSAM use for some individuals.⁸⁸

- Researchers investigating the histories of child pornography^{xiii} offenders found that 63% of contact offenders (i.e., those who had also engaged in physical sexual abuse of minors) and 42% of non-contact offenders (exclusively online offenders) traded **adult** pornography online. **In addition, 40% of contact offenders and 38% of non-contact offenders paid for adult pornography online.** Contact offenders spent more time per week (12 hours per week) than non-contact offenders (7 hours per week) viewing **adult** pornography and pornography in general (24 hours per week vs 18 hours per week). Contact offenders were more likely to use the internet to locate potential sexual abuse victims **and to engage in grooming by sending both child and adult pornography to potential victims.**⁸⁹
- An analysis of data (i.e., images and metadata) extracted from the hard drives of 40 individuals convicted of possession of child pornography between 2004 and 2012 revealed that **the most prevalent pattern was a progressive decrease in the age of the person depicted and a progressive increase in the severity of the sexual acts. All the child pornography collections examined also included mainstream pornography content.** The results also indicated that child pornography offenders may have more deviant sexual interests than offline offenders because online offenders are able to explore a wider range of sexual interests.⁹⁰
- In a survey of 434 men involved in online sexual activities (OSA), by far the most common activity was viewing pornography (99%). The same survey found that **49% of men searched for or were involved in OSA, which previously did not interest them or which they previously found disgusting.**⁹¹
- Additionally, a 2014 voxel-based morphometry and fMRI study conducted by researchers at the Max Plank Institute found that **nonaddictive levels of pornography use are characterized by novelty-seeking conditioning and habituation to sexual stimuli, meaning users may need more novel and extreme content over time to achieve the same level of sexual arousal.**⁹²

The potential for desensitization of users to so-called “vanilla,” hardcore pornography so that they seek harder and more extreme material is extremely concerning, considering how many people now consume pornography. The table below provides a window into the staggering amount of pornography consumption occurring around the world. Billions of people are consuming pornography monthly. If only a fraction of these individuals migrates to child

^{xiii} “Child pornography” is the legal term for child sexual abuse material.

sexual abuse material, it will (and, indeed, already has) result in a tsunami of individuals seeking CSAM.

Table 1: World’s Top Pornography Tube Site Platforms by Global Web Traffic⁹³					
November 2023 – January 2024					
	Global Web Traffic Share	Monthly Visits	Monthly Unique Visitors	Combined Rank of Monthly and Unique Visits	Year Founded
XVideos	11 th	2.632 B	346.6 M	9 th	2007
Pornhub	14 th	2.228 B	344.6 M	13 th	2007
XNXX	15 th	2.180 B	273.4 M	10 th	2007
xHamster	24 th	1.398 B	337.6 M	21 st	2007

In summary, evidence firmly indicates that the use of mainstream, hardcore pornography, especially at younger ages, is a risk factor for escalation to the use of CSAM. The use of CSAM is also linked to hands-on offenses against children.⁹⁴ Thus, mainstream hardcore pornography is a significant contributing factor to this massive public health concern.

B. Sexual Violence

An estimated 736 million women have experienced either physical or sexual violence, or both forms of violence, at least once in their lifetimes.⁹⁵ The World Health Organization reports that sexual violence results in a 2.2 times greater risk of mental health and neurological disorders, 1.8 times greater risk of acquiring a communicable disease, 1.5 times greater risk for poor health-related quality of life, 1.5 times greater risk of health risk behaviors (e.g., substance abuse, alcohol abuse),⁹⁶ and 5.2 times greater risk of subsequent violence victimization.⁹⁷ Mainstream, hardcore pornography deepens these problems. As elucidated below, mainstream, hardcore pornography not only depicts sexual violence but teaches that women enjoy it.

- Multiple studies find increased sexual objectification^{xiv} of women by both men and women who consume pornography, be they adults or adolescents.⁹⁸ This is noteworthy because “considerable research has shown that women

^{xiv} “Sexual objectification is characterized by the reduction of the female body to its physical appearance and the treatment of girls and women as bodies, such that their sexualized bodies or body parts are separated out from their personhood and commodified for others’ consumption or regarded as capable of representing the girls and women themselves.” From: Tomi-Ann Roberts, Rachel M. Calogero, and Sarah J. Gervais, “Objectification Theory: Continuing Contributions to Feminist Psychology,” in *APA Handbook on the Psychology of Women Vol. 1 History, Theory, and Battlegrounds*, eds. in chief, Cheryl B. Travis and Jacquelyn W. White (Washington, DC: American Psychological Association, 2018), 249.

portrayed in sexualized ways are denied subjectivity, agency, and qualities of personhood.”⁹⁹ **Dehumanization resulting from sexual objectification has serious implications for sexual violence, including, but not limited to, diminished view of a sexual assault survivor’s suffering, increased victim blaming, and diminished blame of rapists.**¹⁰⁰

- A content analysis of 4009 heterosexual scenes from 3737 randomly selected videos from Pornhub and XVideos found that 45.1% of Pornhub and 35% of XVideos scenes contained at least one act of physical aggression against a woman or a man. **Women were the targets of physically aggressive acts in 96.7% of Pornhub and 96.8% of XVideos scenes containing physical aggression. Women responded with either pleasure or neutrality in 97.4% of Pornhub and 92.7% of XVideos scenes containing sexual aggression.**¹⁰¹
- Researchers analyzed 131,738 pornography titles^{xv} appearing on **the landing pages** of the three most popular online pornography websites in the United Kingdom (**Pornhub, XVideos, and XHamster**) over a period of six months during 2017-2018.¹⁰² The titles were analyzed for keywords indicating the following four broad categories of **sexual violence**: sexual activity between family members, aggression and assault, image-based sexual abuse, and coercive and exploitative sexual activity. **One in every eight (12%) titles described activities that constitute sexual violence.** By category, the most common form of sexual violence described was sexual activity between family members, followed by physical aggression and sexual assault, image-based sexual abuse, and coercive and exploitative sexual activity, respectively. It is important to recognize that children who access these sites are being exposed and socialized to sexual violence right from the start. Children using these sites are exposed to material served up to cater to a seasoned adult audience.
- A 2022 survey found that 52% of 13- to 17-year-olds who had been exposed to pornography (73% of 1300+ sample) saw violent pornography, including choking (36%), someone in pain (37%), or depictions of what appeared to be rape (19%).¹⁰³ **Longitudinal research shows that childhood exposure to violent pornography predicts a nearly six-fold increase in self-reported sexually aggressive behavior later in life.**¹⁰⁴
- A meta-analysis of 22 studies (cross-sectional and longitudinal) from seven countries documented increased physical and verbal aggression by both male and female pornography consumers, with consumption of violent pornography being an exacerbating factor.¹⁰⁵

^{xv} The original sample included 150,000 titles. Titles not communicating a sexual script (e.g., “share my vid please,” or with only the performers name or name of a production studio) were removed.

- Another meta-analysis of 59 studies of male adolescent sexual offending^{xvi} found increased perpetration of sexual offenses by minors exposed to pornography.¹⁰⁶
- A national longitudinal study of sexual offenses (i.e., sexual harassment, coercive sex, and rape) among youth aged 10-21 years reported that, after controlling for potentially influential characteristics, current exposure to violent pornography was strongly associated with sexual violence perpetration.¹⁰⁷
- Exposure to pornography among adolescents is also linked to their physical and sexual victimization.¹⁰⁸ **Exposure to violent or rape pornography increases a child's odds of experiencing sexual exploitation by nearly three times, as demonstrated by a meta-analysis of 37 studies, including a total of 67,453 participants.**¹⁰⁹

Thus, mainstream hardcore pornography plays a significant role in fueling the public health issue of sexual violence.

C. Sexually Transmitted Infections

We could put forward a variety of sexual behaviors portrayed in pornography that are known risk factors for acquiring sexually transmitted infections, but here we will address only the role of condom use and HIV. **HIV has claimed the lives of an estimated 32.9 to 51.5 million people.**¹¹⁰ There is no cure for those who become infected.¹¹¹ It is also estimated that the promotion and consequent increase in condom use to abate its spread has resulted in about 117 million fewer HIV infections.¹¹²

Ironically, condom use in the pornography industry is miniscule. Multiple content analyses of pornography have found that condom use ranges from 2 to 11%, and in one case, only 1 out of 45 videos.¹¹³ As one performer put it, "You either accept the fact that you're going to do a scene without a condom, or accept the fact that you're not doing a scene."¹¹⁴

In addition to the risk this poses to the health of pornography performers, the concern arises that consumers of pornography are socialized to reject condom use "as condomless sex is a rewarded behavior in pornography."¹¹⁵ This concern has been affirmed by **a meta-analysis including data from 18 countries and more than 35,000 participants, which found higher pornography consumption was associated with a higher likelihood of engaging in condomless sex.**¹¹⁶ Subsequent studies also support this finding¹¹⁷ and the association increases the more one perceives pornography as realistic.¹¹⁸

^{xvi} The review included 3,855 male adolescents who had sexually offended; 13,393 non-offending male adolescents.

D. Mental Health

According to recent data from Mental Health America (MHA), **21% of adults are experiencing mental illness—more than 50 million Americans.**¹¹⁹ MHA also reports that 16% of youth (aged 5 – 21) experienced a major depressive episode in the last year, and that **11.5% of youth, more than 2.7 million, are experiencing severe major depression.**¹²⁰ Further, “In the 10 years leading up to the [COVID-19] pandemic, feelings of persistent sadness and hopelessness—as well as suicidal thoughts and behaviors—increased by about 40% among young people, according to the Centers for Disease Control and Prevention’s (CDC) Youth Risk Behavior Surveillance System.”¹²¹ While many factors are contributing to this situation, once again, mainstream, hardcore pornography is a risk factor.

- An Australian study of 941 persons aged 15 to 29 years found that 87% had ever viewed pornography. Compared to those with no recent mental health problems, **those with recent mental health problems were more likely to be moderate to frequent pornography users.** Those who first started viewing pornography at a younger age were also more likely to report experiencing a mental health problem within the past six months.¹²²
- In a cross-sectional study of 792 emerging adults from college campuses, researchers observed that **higher pornography use was significantly associated with less self-worth and more depressive symptoms.** This was consistent after controlling for age, religiosity, impulsivity, race, and parent’s marital status.¹²³
- Analysis of online survey data obtained from 1,539 German-speaking pornography users (73% male; 27% female) between 18 and 76 years old found that 5.9% (7.15% men; 2.61% women) reached a score on the Online Pornography Disorder Questionnaire indicating problematic pornography use.¹²⁴ **Participants with problematic pornography use scored significantly worse in every measure of psychological functioning considered, including somatization, obsessive-compulsive behavior, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism.** Furthermore, most results were elevated to a clinically relevant degree when compared to the general population. **The study authors characterized the intensity of the problems experienced by problematic pornography users as “severe psychological distress.”**¹²⁵
- In a study utilizing survey data of 1,501 children and adolescents (ages 10–17 years), questions based on the nine criteria for major depression in the DSM IV were used to assess for depressive symptoms. Youth reporting five or more clinical features were categorized as reporting a high level of depression. The study reported that clinical features of major depression were more than two times greater among children and adolescents who reported seeking online

pornography. Specifically, **11% of youth who sought online pornography reported major depression compared to 4% among those who sought pornography offline and 5% who did not.**¹²⁶

E. Addiction-related Brain Changes

Finally, we come to the robust evidence that pornography induces addiction-related changes to the brain's structure and function. While there are more than sixty neurological studies that support the view that pornography consumption may result in behavioral addiction, and none to our knowledge falsify this claim, we provide findings from three notable studies on this topic.

- **Decreased Brain Matter in the Right Caudate of the Caudate Nucleus:** A 2014 fMRI study of 64 pornography users found that increased pornography use (i.e. pornography dosage) is **linked to decreased grey matter** in brain regions associated with motivation. **A higher dosage of pornography consumption among even non-addicted, moderate pornography consumers was related to both greater gray matter volume loss and frontal connectivity deterioration.** These changes involve regions of the brain associated with substance abuse disorders and may reflect changes in neural plasticity as a result of acute stimulation of the reward system.¹²⁷
- **Novelty-Seeking and Conditioning:** Using fMRI, a 2015 study from Cambridge compared the brains of those with compulsive sexual behavior to healthy volunteers. Researchers found that compulsive sexual behavior is characterized by novelty-seeking, conditioning, and habituation to sexual stimuli in males—meaning users need more extreme content over time in order to achieve the same level of arousal.¹²⁸
- **Disassociation between “Wanting” and “Liking”:** this double-blind, experimental study applied a popular theory of addiction to pornography use by comparing the brain responses of men with and without problematic pornography use (PPU).^{xvii} Men with PPU and compared to control subjects displayed similar reaction times to cues predicting monetary reward, but not cues predicting sexual reward. The shortest reaction times to sexual reward cues were observed among PPU subjects. The “bigger” the anticipated reward, the quicker their reaction times. This demonstrates their higher motivation to view sexual rewards than control subjects and provides support for incentive sensitization theory rather than general impairments of reward processing. Additionally, brain responses to sexual pictures in PPU men revealed a dissociation between measures of “wanting” and “liking” the sexual imagery but not in men without PPU, providing strong evidence that PPU may represent a behavioral addiction. This dissociation was significantly

^{xvii} Men seeking treatment for PPU met criteria of “hypersexual disorder” and were without other psychiatric diagnoses.

correlated to the severity of the PPU subjects' compulsive sexual behaviors, the amount of pornography consumed, and the frequency of masturbation reported.¹²⁹

CONCLUSION

More than 50 years ago, the Supreme Court of the United States recognized that there were multiple governmental interests in prohibiting obscenity (i.e., hardcore pornography). In *Paris Adult Theatre I v. Slaton*, 413 U.S. 49 (1973), the Court stated, “In particular, **we hold that there are legitimate state interests at stake in stemming the tide of commercialized obscenity**, even if it is feasible to enforce effective safeguards against exposure to juveniles and to passersby . . . **These include the interest of the public in the quality of life and total community environment, the tone of commerce . . . and, possibly, the public safety itself** (57-58).”¹³⁰ (emphasis added)

The developments of the past five decades soundly attest that it was right.

Allowing AGI to further fuel child sexual abuse, image-based sexual abuse, and other forms of sexual abuse, violence, and exploitation is a grave misuse of the power and promise of AGI. In view of the U.S. Supreme Court and the body of evidence provided above, **the National Center on Sexual Exploitation calls on OpenAI to recognize that its paramount responsibility is to safeguard the wellbeing of humanity from the unintended consequences of AGI and to rectify harms AGI has already thrust upon us.** Specifically, OpenAI must acknowledge:

- its responsibility as a frontier AI company to maintain the highest ethical standards and to protect human life and wellbeing
- that failure to prioritize safety within the AI sector has already unleashed a tsunami of sexual harm upon thousands of individuals, the repercussions of which will last their lifetimes
- the necessity of ensuring that AI pre-training and training data sets are thoroughly purged of *all* sexually explicit and exploitative material (i.e., hardcore pornography depicting adults, CSAM, IBSA, and AI-generated versions of such material)
- that failure to remove all sexually explicit and exploitative material results in abuse-trained models that normalize sexual exploitation and further exploit the exploited
- that OpenAI’s partnership with Reddit is posed to replicate mistakes akin to those of Stability AI’s LAION 5b text-image model unless *very robust* measures are taken to filter out the massive quantities of hardcore pornography and sexual exploitation material in its data
- that broad harms to public health are incurred and intensified by the widespread availability of hardcore pornography (i.e., “NSFW” material) on the Internet, and
- the unqualified impossibility of the “ethical” generation of “NSFW” sexually explicit material by AGI.

OpenAI must also take the following immediate actions:

- reestablish and empower its “Superalignment” group to prioritize safety and adherence to ethical practices
- publicly release data regarding the provenance of its pre-training and training datasets and detailing its efforts to remove sexually explicit material
- fix the deficiencies of CLIP and work with users of CLIP to ensure that resulting datasets are free from all sexually explicit material
- launch robust efforts to ensure that all its existing pre-training and training datasets are thoroughly purged of hardcore pornography, CSAM, IBSA, as well as AI-generated versions of these materials
- move immediately to ensure that datasets obtained from Reddit are robustly vetted and purged of all hardcore pornography and sexual exploitation material
- review and update its Usage Policy and Approach to Frontier Risk to address the deficiencies raised above
- quickly announce a decision to **not** allow the use of any OpenAI model or tool for the generation of “NSFW” material.

A range of appalling sexual harms resulting from AGI are happening now. Failure to take the steps outlined above will allow these harms to proliferate, posing severe and intolerable risks to humanity. Thus, NCOSE, survivors, our allies, and all who hope for a world free from sexual abuse and exploitation look expectantly for OpenAI to act decisively to implement these and any other measures within its powers, to mitigate future risks, ameliorate current impacts, and stop further sexual abuse and exploitation facilitated by AGI.

APPENDIX A

National Center on Sexual Exploitation Dirty Dozen List 2024: Reddit Notification Letter



Established 1962

KINDSEY PENTECOST CHADWICK | Interim President

DAWN HAWKINS | CEO

BENJAMIN A. BULL, ESQ. | SVP & General Counsel

BOARD MEMBERS

RON DEHAAS | Chairman
Founder & CEO, Covenant Eyes

DAN O'BRYAN, ESQ. | Vice Chairman
Former Fellow, Weatherhead Center for International Affairs, Harvard University

RHONDA GRAFF | Secretary
Former Aerospace Program Manager/Industrial Engineer

BETH COONS | Treasurer
Chairman, Fairwold, Companies

NATHAN ANDERSEN, ESQ.
General Counsel, LeSeuer Investments

KINDSEY PENTECOST CHADWICK | Interim President
President Emeritus, ItWorked! International

JOHN FOUBERT, PH.D.
Dean, College of Education, Union University

ROBERT GEORGE, PH.D.
McConnick Professor of Jurisprudence,
Purdue University

HON. TANYA GOULD
Director, Virginia Attorney General's Office on
Human Trafficking

DONALD L. HILTON, M.D.
Neurosurgeon, University of Texas - San Antonio

DONNA M. HUGHES, PH.D.
Retired Professor, Cation Endowed Chair,
University of Rhode Island

AMBASSADOR SWANEE HUNT
Chair and Founder, Swanee Hunt Alternatives;
Eleanor Roosevelt Lecturer of Public Policy,
Harvard Kennedy School of Government

MARY ANNE LAYDEN, PH.D.
Director, Sexual Trauma and Psychopathy Program,
University of Pennsylvania

MOST REVEREND PAUL S. LOVERDE
Bishop, Diocese of Arlington, VA, Retired

SHERRY NILLES
National Regent, Catholic Daughters of the Americas

PATRICK A. TRUMAN
President & CEO Emeritus of NCOSE

LINNEA W. SMITH, M.D.
Retired, Psychiatrist

MELEA STEPHENS, L.P.C., M.M.F.T.
Marriage and Family Therapist

JASON SYVERSON
Chair & Chief Investment Officer, SS
Impact Foundation

CAROLYN WEST, PH.D.
Professor, Clinical Psychology,
University of Washington

April 8, 2024

Mr. Steve Huffman
Co-Founder and CEO, Reddit
420 Taylor Street
San Francisco, CA 94102

Re: Reddit's placement on the 2024 Dirty Dozen List for hosting image-based sexual abuse, child sex abuse material, and promoting violence against women

Dear Mr. Huffman:

We write to inform you that **Reddit will once again be named, for the fourth consecutive year, to the annual Dirty Dozen List featuring 12 mainstream contributors to sexual exploitation and abuse.** The National Center on Sexual Exploitation—a non-partisan nonprofit based in Washington, DC, dedicated to addressing the full spectrum of sexual exploitation—calls on Reddit to urgently address the extensive harms and crimes that thrive on your platform. This year's Dirty Dozen List will be revealed on Wednesday, April 10, 2024.

Reddit is known as one of the most popular discussion platforms on the Internet right now, with about [73 million active users each day](#).¹ In 2023, Reddit averaged [1.2 million posts and 7.5 million comments daily](#).² And in just six months, from January to June 2023, Reddit users [exchanged 472,984,011 personal messages](#).³

Since its inception, Reddit has been known as a hub for multiple forms of sexual abuse and exploitation: child sex abuse material posting and trading, image-based sexual abuse, sex trafficking, and child sexual exploitation. Reddit has continued to allow the proliferation of this content and behavior despite its public attempts at cleaning up its image ahead of going public on March 21, 2024.

The following reports from researchers and online safety organizations demonstrate the wide-ranging sexual exploitation occurring across Reddit:

⇒ The number of links advertising “undressing apps” [increased more than 2,400% across social media platforms like Reddit and X in 2023](#) (Graphika, 2024).⁴

1201 F St. NW, Suite 200, Washington, DC 20004 | 202.393.7245 | EndSexualExploitation.org | public@ncose.com

BUILDING A WORLD WHERE ALL PEOPLE CAN LIVE AND LOVE FREE FROM SEXUAL ABUSE AND EXPLOITATION

- ⇒ [#3 app flagged for severe sexual content](#) (Bark, 2023).⁵
- ⇒ Reddit is [one of the top eight social media platforms](#) used to search, view, and share child sexual abuse material (CSAM) and top eight social media platforms used to establish contact with a child (Suojellaan Lapsia, Protect Children, February 2024).⁶
- ⇒ Percentage of minors with self-generated child sexual abuse material ([SG-CSAM](#)) [experiences that use Reddit at least once per day](#): 34% of minors who have shared their own SG-CSAM; 26% of minors who have reshared SG-CSAM of others; 23% of minors who have been shown nonconsensually reshared SG-CSAM (Thorn, November 2023).⁷
- ⇒ [One of the top ten online platforms where teens report having seen pornography](#) (UK Children’s Commissioner, 2023).⁸
- ⇒ Australian eSafety Commissioner research identified [Reddit as one of the most popular social media](#) sites where young people report seeing pornography (Australian eSafety Commissioner, 2023).⁹

Reddit’s recent [partnership with Google](#), disclosed in the March 2024 S-1 filing with the SEC, adds another layer of concern.¹⁰ This deal, which is expected to generate \$66.4 million in revenue by the end of 2024, [allows Google to train AI models using Reddit’s content](#).¹¹ This arrangement is particularly alarming, given Reddit’s historical challenges with proactively identifying and eliminating sexually explicit content, including abusive and illegal imagery. Despite new policies banning [third party and API access to sexually explicit content](#),¹² without stringent safeguards to filter out image-based sexual abuse and CSAM, the **potential for Google’s AI models to be trained on the pervasive amount of sexually exploitative and potentially illegal material on Reddit is much too high.**

NCOSE finds this move incredibly concerning, especially considering Reddit’s lack of effective content blurring measures and the absence of age or consent verification mechanisms. The platform’s approach to moderation and its partnership with Google highlight profound shortcomings. These issues not only jeopardize efforts to ensure online safety but also position Reddit as a platform that facilitates the spread of sexual exploitation and abuse, raising serious questions about its [newly proclaimed](#) viability as a ‘safe and responsible’ digital space.¹³

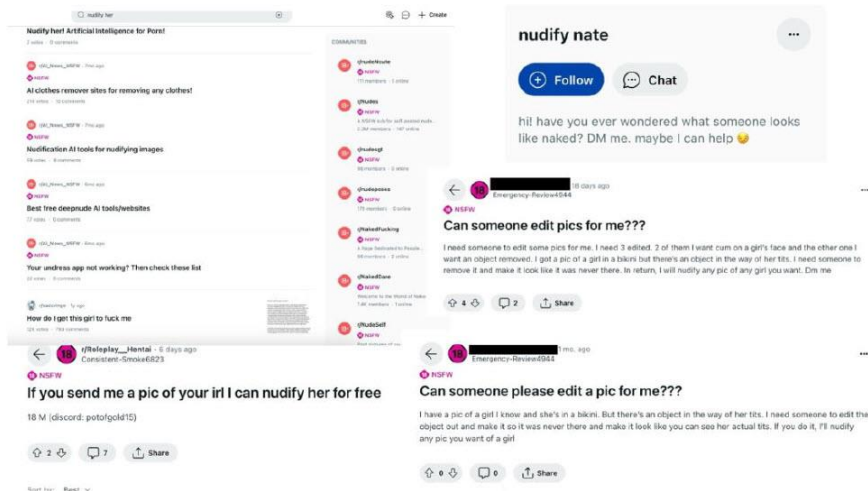
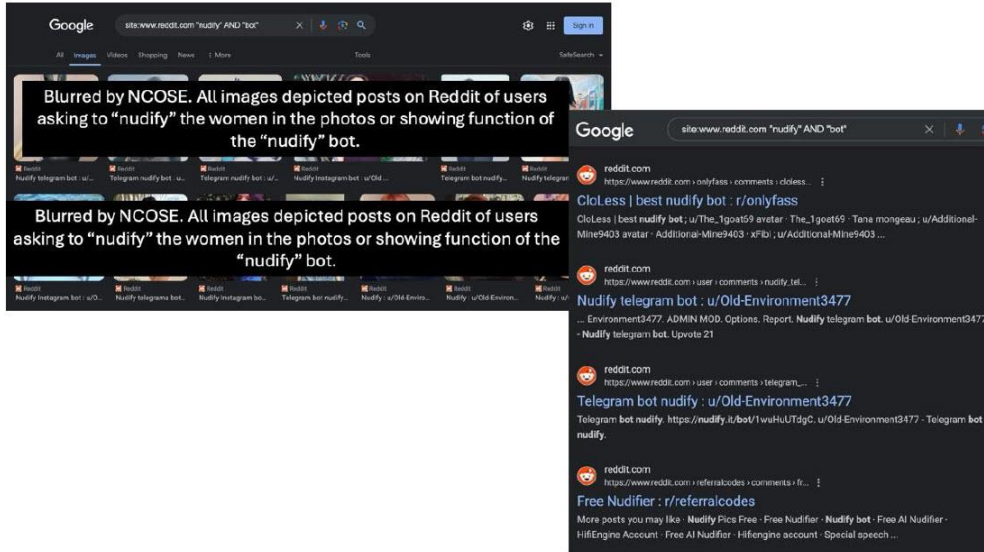
We do not need to reiterate the evidence in detail, as you have been notified many times, and we feel confident that you are well aware of the dangers occurring on your platform. Until you radically reform your current policies to ensure that reports are acted on, and that age and consent verification is implemented, NCOSE will continue elevating the destructive realities of Reddit’s platform to law enforcement, policymakers, the press, and parents. Most importantly, we will continue to remind your potential investors that Reddit is a dangerous and risky investment.

Image-Based Sexual Abuse and AI-Facilitated Sexual Abuse and Exploitation

Despite Reddit’s update to [Rule 3](#) in June 2023,¹⁴ the new policy literally titled, “[Never Post Intimate or Sexually Explicit Media of Someone Without Their Consent](#),”¹⁵ which aimed to include deepfakes and AI-generated explicit content featuring identifiable individuals, a critical gap remains: Reddit’s failure to tackle AI-generated pornography of adults, unless it depicts identifiable persons, starkly overlooks the fundamental issue of consent. Reddit’s historical negligence of content moderation and permissiveness for [gray area content](#)¹⁶ has allowed exploitative and potentially illicit content to proliferate.

For example, December 2023 research found a [2,400% surge in ads for “undressing apps.”](#)¹⁷ despite Reddit’s [policies against non-consensual content](#)¹⁸ and [sexually explicit ads.](#)¹⁹ underscoring the platform’s inadequacies in curbing image-based sexual abuse.

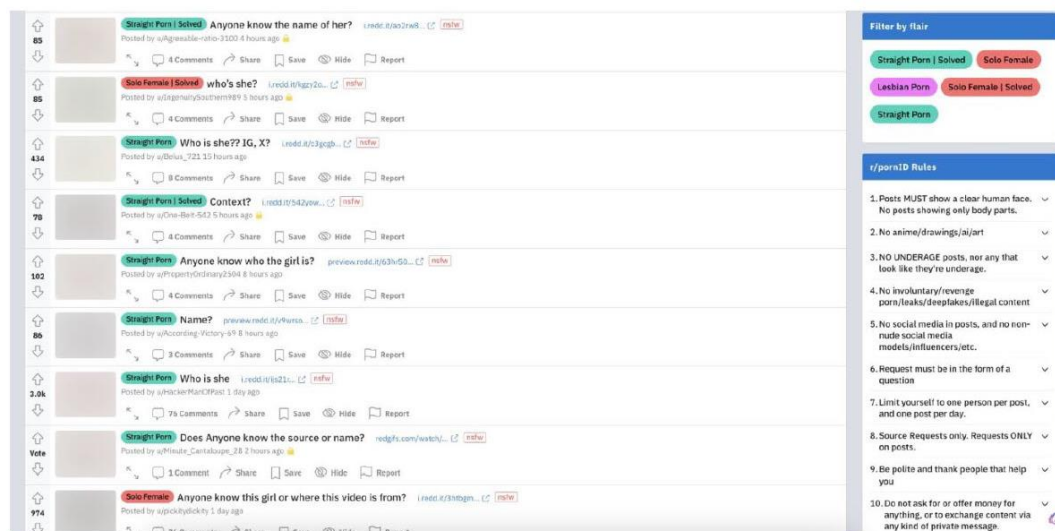
Reddit has also failed to properly moderate content depicting, soliciting, or offering to create “deepfake” pornography or “nudified” images of private individuals and celebrities. In fact, many posts on Reddit even monetize creating sexual “deepfake” and “nudified” images of any person someone submits.



Reddit also has a number of communities dedicated to the creation and distribution of AI-generated pornography. A dedicated AI-generated pornography community NCOSE reported on in 2023 [grew by almost 3x the number of subscribers in a year from 48,000 to 128,000 members](#).²⁰ This exploitative tool amassed over 15 million “donated” sexually explicit and pornographic images to train their technology, and posts AI-generated pornography by the hour on Reddit. During its soft launch in July 2023, the Unstable Diffusion CEO and Co-Founder stated the [model was generating over 500,000 images a day](#).²¹ A 2023 Forbes article also highlighted the [absolute explosion of AI-generated pornography on Reddit](#),²² indicating that the platform’s ambiguous policies have allowed it to flourish.

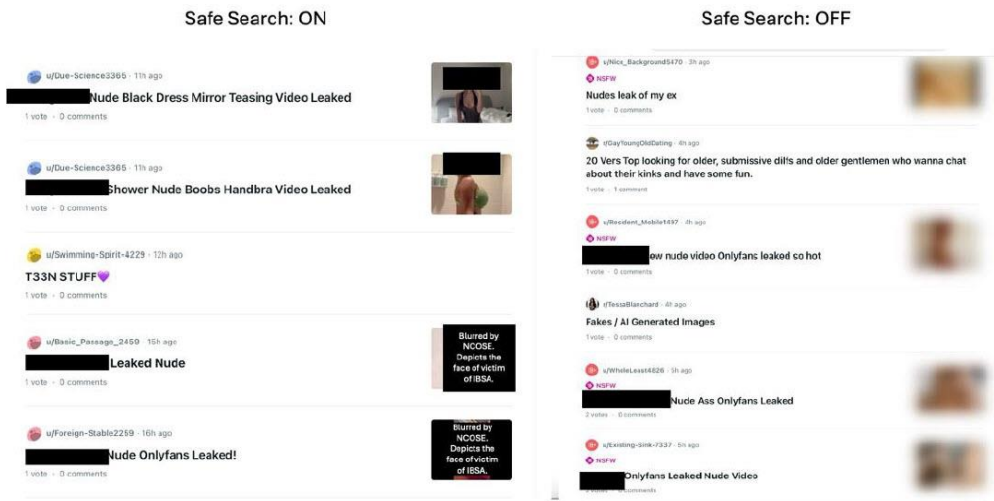
Although Reddit has improved blurring mechanisms for sexually explicit content, **blurring non-consensual sexually explicit imagery does not make the content any less “non-consensual.”** This feature is *easily* bypassed, allowing Redditors to simply click on the blurred post to view the image. NCOSE is deeply concerned by the growth of subreddits dedicated to non-consensual sharing of sexually explicit images, not only persisting but growing.

The subreddit r/PornID (dedicated to identifying women in sexually explicit imagery and videos) increased in subscribers from **1.3 million in 2023 to 1.7 million in 2024**. Although the thumbnails are blurred, this does not negate the fact that this subreddit, like dozens of others NCOSE researchers found, is dedicated to sharing non-consensual content. Further, this subreddit encourages posting of non-consensual content for the purpose of identifying the women in such content, potentially resulting in other forms of abuse such as doxing, stalking, and harassment.



The fact that NCOSE researchers *easily* found dozens of subreddits containing language indicating nonconsensual content as the main focus, with blurred and unblurred imagery, is also indicative of Reddit’s failure in enforcing its newly updated policies.

Searches with Reddit’s “Safe Search” Content Filter ON returned posts containing non-consensual content or links to such content. Further, when this filter is ON, sexually explicit content is openly shown as the blurring filter only functions when Safe Search is OFF. For example, the search “leaked nudes” returned literally hundreds of posts from Redditors soliciting or offering leaked sexually explicit imagery of women from OnlyFans, ex-partners, “snapchatsluts,” and more in the last year alone. Searches for downblousing similarly returned an alarming number of results despite this being a known IBSA term and prominent form of abuse.



Considering Reddit’s historic negligence in moderating sexually explicit content, it is deeply concerning that this content is still slipping through Reddit’s ‘improved’ filters.

Child Sexual Abuse Material and Sexualization of Minors

NCOSE [acknowledges](#) that in June 2023,²³ Reddit took a step forward by announcing [updates to its Child Safety Policy](#)²⁴ (Rule 4),²⁵ including comprehensive measures against the sexualization of children – something NCOSE has found to be pervasive across the platform. These measures ranged from banning suggestive images of children, including those that are AI-generated, to prohibiting behaviors that exploit minors.

However, despite these updated regulations and Reddit’s efforts to enhance moderation, tangible enforcement has been lacking. Recent research underscores the persistent availability of CSAM, content that sexualizes minors, and AI-generated sexually explicit and suggestive content of minors on Reddit. For example, research conducted by the Stanford Internet Observatory in December 2023 [found child sexual abuse material \(CSAM\) originating on Reddit](#) had been used to train AI models such as Stable Diffusion.²⁶ Further, the actions taken by the Australian eSafety Commissioner – one of the leading

authorities on online child safety in the world – who [issued a legal notice to Reddit](#) (amongst five other platforms) requesting a detailed inquiry into Reddit’s practices to detect and remove CSAM from the platform, suggests additional concern.²⁷

Unfortunately, we have yet to see substantial evidence of comprehensive enforcement since Reddit updated their child safety policy:

- With Safe Search **ON**, NCOSE researchers found that the search “mega teen leaks” and “mega links” **returned over 100 results in the last 60 days**, most openly soliciting or searching for links demonstrating high indication of child sexual abuse material (CSAM). (NCOSE, April 2024).
- Content attached to posts from the above search results contained imagery suggesting child sexual abuse between adults and minors. One post titled, “MEGA links telegram::@cum_file5,” contained an image with the text, “Boys for Sale,” depicting a fully clothed adult man covering the mouth and grabbing the genital region of an underage boy, who was wearing only his underwear and had visible red marks of abuse across his chest (NCOSE, April 2024).
- NCOSE researchers conducted a Google search for “site:www.reddit.com ‘mega’ AND ‘teen’” and received **14,500 results of content on Reddit containing high indications of child sexual abuse imagery sharing, trading, and selling** (NCOSE, April 2024).
- NCOSE researchers analyzed the top 100 pornography subreddits (what is tagged on Reddit as “NSFW”) and found numerous examples of child sexualization and/or possible indicators of child sexual abuse material—including titles like “xsmallgirls” and references to flat-chestedness (NCOSE, July 2023).

Search results conducted on April 5, 2024, for “site:www.reddit.com ‘mega’ AND ‘teen’” returned 14,500 results of extremely concerning content.



In addition to this, NCOSE researchers have found artificial / illustrated pornography depicting bestiality and even illustrated bestiality involving child-like figures.

In 2023, NCOSE exposed a subreddit dedicated to AI-generated bestiality imagery, some of which depicted child-like figures. While the original subreddit we reported was removed, the owner *EASILY* made a new account and subreddit with the same title – and even admitted to doing so (proof available upon request). The new subreddit is still active as of April 5, 2024, and contains imagery depicting bestiality involving child-like figures and posts of users sharing their experiences of *actual* bestiality. It is horrendous that Reddit allows this clearly nefarious and potentially illegal content to remain.

It appears that Reddit knowingly allows CSAM, sex trafficking, and pedophilic activity to thrive on its platform. Below are details of just a few cases in 2023 and 2024:

- On February 2, 2024, an [Ashland, KY man was charged](#) with the sharing and distribution of nude images of young boys aged 10-14. He had distributed them to as many as 20 others on Reddit.²⁸
- On January 25, 2024, a 37-year-old [Dover, NJ man was arrested for possessing CSAM](#) and uploading it onto Reddit. The initial report was made following a NCMEC CyberTip, as reported to the CyberTipline.²⁹
- In November 2023, [a former associate pastor of the Baptist church and Chaplain at the Oklahoma State Penitentiary was arrested](#) for engaging in sexually explicit chats with minors on Reddit and possessing CSAM.³⁰
- On August 30, 2023, an [Albert Lea, Minnesota man pleaded guilty to four counts of possession of CSAM](#) after admitting to buying CSAM off of Reddit. He was arrested based on a CyberTip from NCMEC.³¹
- In August 2023, a [Glendale, Wisconsin police officer was fired and arrested for posting CSAM on Reddit](#). He had at least six images in his possession at the time of the arrest.³²

Such findings point to a troubling disconnect between Reddit’s stated policies and the platform’s ability to enforce them effectively. The continued presence of subreddits that sexualize minors, found even with the Sensitive Content filter enabled, starkly illustrates the challenges Reddit faces in its quest to safeguard against the exploitation of minors, casting doubt on the platform’s ability to provide a safe and secure online environment.

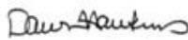
Reddit is dangerous by design. For almost two decades, Reddit has failed to implement effective and meaningful policies, touting free speech and privacy over the safety and well-being of its users – despite the overwhelming evidence of harmful and violent behavior facilitated on your platform.

We urge Reddit to implement the following recommendations with the urgency commensurate with the gravity of harm on your platform:

1. **Implement robust age and consent verification measures** – ban pornography and sexually explicit content until such measures are **actively enforced** to prevent image-based sexual abuse and child sexual abuse content on Reddit.
2. **Require mandatory email verification for current and future users, as anonymity allows for abuse without accountability.**

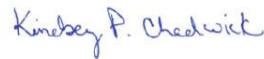
3. **Make Reddit 17+ and prevent minors from accessing your platform** – both Apple App Store and Google Play age ratings for Reddit are 17+ and M for “Mature,” respectively. *Given that pornography is so easily accessible and other harms are rampant, children should not be on Reddit; asking Reddit for parental controls or other safety measures would be like asking Pornhub to make itself kid-friendly.*
4. **Ban all forms of AI-generated pornography on Reddit** – Reddit has no way of confirming age and consent verification of images used in training sets or the individuals depicted in generated images. Reddit’s June 2023 clarification to only ban AI-generated pornography of identifiable adults is another example of purposefully creating ambiguous policies providing ample space for exploitation and loopholes of such policies.
5. **Ban external links to pornography and sexually explicit media** – the links NCOSE researchers found frequently led to sexually explicit content on pornography sites and social media platforms that contained indicators of IBSA and CSAM.

Thank you for your attention,



Dawn Hawkins

CEO



Kindsey P. Chadwick

Interim President

Cc:

Jen Wong, Chief Operating Officer

Chris Slowe, Chief Technology Officer

Roxy Young, Chief Marketing and Consumer Experience Officer

Pali Bhat, Chief Product Officer

Shariq Rizvi, Executive Vice President of Ads Monetization

¹ Reddit, Inc., “Amendment No. 3 to Form S-1 Registration Statement,” United States Securities and Exchange Commission, March 19, 2024, <https://www.sec.gov/Archives/edgar/data/1713445/000162828024011789/reddit-sx1a3.htm>.

² Ibid.

³ “Transparency Report: January to June 2023,” Reddit, accessed April 8, 2024, <https://www.redditinc.com/policies/2023-h1-transparency-report>.

⁴ Santiago Lakatos, “A Revealing Picture: AI-Generated ‘Undressing’ Images Move from Niche Pornography Discussion Forums to a Scaled and Monetized Online Business,” *Graphika*, December 8, 2023, <https://graphika.com/reports/a-revealing-picture>.

⁵ Bark, “What Being Online Was Like for Kids in 2023,” Bark’s 2023 Annual Report, accessed April 8, 2024, <https://www.bark.us/annual-report-2023>.

ENDNOTES

- ¹ OpenAI.com, “About,” n.d., <https://openai.com/about/> (accessed May 16, 2024).
- ² OpenAI.com, “Our Approach to Alignment Research,” August 24, 2022, <https://openai.com/index/our-approach-to-alignment-research/> (accessed May 15, 2024).
- ³ OpenAI.com, “Democratic Inputs to AI Grant Program: Lessons Learned and Implementation Plans,” January 16, 2024, <https://openai.com/index/democratic-inputs-to-ai-grant-program-update/> (accessed May 15, 2024).
- ⁴ Associated Press, “Seoul AI Summit Opens with Companies Including Google, Meta, and OpenAI Pledging to Develop AI Safely,” Mynews13.com, May 21, 2024, <https://mynews13.com/fl/orlando/ap-top-news/2024/05/21/seoul-ai-summit-opens-with-companies-including-google-meta-openai-pledging-to-develop-ai-safely> (accessed May 22, 2024).
- ⁵ Thorn, “Thorn and All Tech is Human Forge Generative AI Principles with AI Leaders to Enact Strong Child Safety Commitment,” Thorn.org, April 23, 2024, <https://www.thorn.org/blog/generative-ai-principles/> (accessed May 23, 2024).
- ⁶ OpenAI.com, “Model Spec,” May 8, 2024, <https://cdn.openai.com/spec/model-spec-2024-05-08.html> (accessed July 23, 2024).
- ⁷ OpenAI.com, “Model Spec,” May 8, 2024, <https://cdn.openai.com/spec/model-spec-2024-05-08.html> (accessed July 23, 2024).
- ⁸ Shayne Longpre, Robert Mahari, Anthony Chen, et al., “The Data Provenance Initiative: A Large Scale Audit of Dataset Licensing & Attribution in AI,” *arXiv* (2023), <https://doi.org/10.48550/arXiv.2310.16787>; Admin Staff, “What DALL-E Dataset Did OpenAI Use?” NightCafe, June 15, 2022, <https://nightcafe.studio/blogs/info/what-dall-e-dataset-did-openai-use> (accessed May 15, 2024).
- ⁹ Vinay Uday Prabhu and Abeba Birhane, “Large Datasets: A Pyrrhic Win for Computer Vision?” *IEEE Winter Conference on Applications of Computer Vision (WACV)* (2020): 1536–1546, <https://doi.org/10.1109/WACV48630.2021.00158>.
- ¹⁰ Alex Fang, Simon Kornblith, and Ludwig Schmidt, “Does Progress on ImageNet Transfer to Real-world Datasets?” 37th Conference on Neural Information Processing Systems (NeurIPS 2023) Track on Datasets and Benchmarks, https://proceedings.neurips.cc/paper_files/paper/2023/hash/4eb33c53ed5b14ce9028309431f565cc-Abstract-Datasets_and_Benchmarks.html (accessed May 16, 2023).
- ¹¹ Vinay Uday Prabhu and Abeba Birhane, “Large Datasets: A Pyrrhic Win for Computer Vision?” *IEEE Winter Conference on Applications of Computer Vision (WACV)* (2020): 1536–1546, <https://doi.org/10.1109/WACV48630.2021.00158>.
- ¹² OpenAI.com, “DALL-E 2 Pre-training Mitigations,” June 28, 2022, <https://openai.com/index/dall-e-2-pre-training-mitigations/> (accessed May 14, 2024).
- ¹³ OpenAI, “DALL-E 2, n.d., <https://openai.com/index/dall-e-2/> (accessed May 15, 2024).
- ¹⁴ Common Sense Media, “AI Ratings DALL-E,” October 27, 2023, <https://www.common Sense Media.org/ai-ratings/dall-e> (accessed July 23, 2024).
- ¹⁵ Ibid.
- ¹⁶ Abeba Birhane, Vinay Uday Prabhu, and Emmanuel Kahembwe, “Multimodal Datasets: Misogyny, Pornography, and Malignant Stereotypes,” *arXiv* abs/2110.01963 (2021): <https://arxiv.labs.arxiv.org/html/2110.01963>.
- ¹⁷ U.S. Department of Justice, Federal Bureau of Investigations, *Uniform Crime Report, Crime in the United States, 2017, Rape*, (2018) <https://ucr.fbi.gov/crime-in-the-u.s/2017/crime-in-the-u.s.-2017/topic-pages/rape> (accessed May 21, 2024).
- ¹⁸ Alexandra (Sasha) Luccioni and Joseph D. Viviano, “What’s in the Box? A Preliminary Analysis of Undesirable Content in the Common Crawl Corpus,” *arXiv* (2021), <https://doi.org/10.48550/arXiv.2105.02732>.
- ¹⁹ Alexandra (Sasha) Luccioni and Joseph D. Viviano, “What’s in the Box? A Preliminary Analysis of Undesirable Content in the Common Crawl Corpus,” *arXiv* (2021), <https://doi.org/10.48550/arXiv.2105.02732>.

-
- ²⁰ LAION.AI, "LAION," <https://laion.ai> (accessed May 16, 2024).
- ²¹ David Thiel, *Identifying and Eliminating CSAM in Generative ML Training Data and Models*, (Stanford Internet Observatory, Cyber Policy Center, December 23, 2023), 3, https://stacks.stanford.edu/file/druid:kh752sm9123/ml_training_data_csam_report-2023-12-23.pdf (accessed May 16, 2024).
- ²² David Thiel, *Identifying and Eliminating CSAM in Generative ML Training Data and Models*, (Stanford Internet Observatory, Cyber Policy Center, December 23, 2023), 7, https://stacks.stanford.edu/file/druid:kh752sm9123/ml_training_data_csam_report-2023-12-23.pdf (accessed May 16, 2024).
- ²³ David Thiel, *Identifying and Eliminating CSAM in Generative ML Training Data and Models*, (Stanford, CA: Stanford Internet Observatory, Cyber Policy Center, December 23, 2023), 11, https://stacks.stanford.edu/file/druid:kh752sm9123/ml_training_data_csam_report-2023-12-23.pdf (accessed May 16, 2024).
- ²⁴ Alexandra S. Levine, "Stable Diffusion 1.5 Was Trained on Illegal Child Sexual Abuse Material, Stanford Study Says," *Forbes*, December 20, 2023, <https://www.forbes.com/sites/alexandralevine/2023/12/20/stable-diffusion-child-sexual-abuse-material-stanford-internet-observatory/?sh=6c134a8f5f21> (accessed May 16, 2024).
- ²⁵ David Thiel, *Identifying and Eliminating CSAM in Generative ML Training Data and Models*, (Stanford, CA: Stanford Internet Observatory, Cyber Policy Center, December 23, 2023), 14, https://stacks.stanford.edu/file/druid:kh752sm9123/ml_training_data_csam_report-2023-12-23.pdf (accessed May 16, 2024).
- ²⁶ Zara Ward, *Revenge Porn Helpline Report 2022*, <https://swgfl.org.uk/assets/documents/rp-report.pdf> (accessed June 13, 2023); Nicholas Kristof, "The Children of Pornhub. Why Does Canada Allow this Company to Profit off Videos of Exploitation and Assault?" *New York Times*, December 4, 2020, <https://www.nytimes.com/2020/12/04/opinion/sunday/pornhub-rape-trafficking.html> (accessed May 16, 2024); Sebastian Meineck and Chris Köver, "World's Largest Porn Site: How XVideos Does Very Little against Sexualized Violence," (*Netzpolitik.org*, March 11, 2022), <https://netzpolitik.org/2022/how-xvideos-does-very-little-against-sexualized-violence/> (accessed May 16, 2024); Sebastian Meineck and Yannah Alfering, "Inside xHamster: Undercover im Löschteam der Meistbesuchten Pornoseite Deutschlands," *Vice*, October 27, 2020, <https://www.vice.com/de/article/akdzdp/inside-xhamster-undercover-im-loeschteam-der-pornoseite> (accessed May 16, 2024).
- ²⁷ Standing Committee on Access to Information, Privacy and Ethics, Canada House of Commons, *Ensuring the Protection of Privacy and Reputation on Platforms Such as Pornhub. Report of the Standing Committee on Access to Information, Privacy and Ethics*, June 2021, 43rd Parliament, 2nd Session, <https://www.ourcommons.ca/Content/Committee/432/ETHI/Reports/RP11148202/ethirp03/ethirp03-e.pdf> (accessed May 22, 2024).
- ²⁸ Nicholas Kristof, "The Children of Pornhub," *The New York Times*, December 4, 2020, <https://www.nytimes.com/2020/12/04/opinion/sunday/pornhub-rape-trafficking.html> (accessed May 22, 2024).
- ²⁹ Standing Committee on Access to Information, Privacy and Ethics, Canada House of Commons, *Ensuring the Protection of Privacy and Reputation on Platforms Such as Pornhub. Report of the Standing Committee on Access to Information, Privacy and Ethics*, June 2021, 43rd Parliament, 2nd Session, 11, <https://www.ourcommons.ca/Content/Committee/432/ETHI/Reports/RP11148202/ethirp03/ethirp03-e.pdf> (accessed May 22, 2024).
- ³⁰ Lisa L. Thompson, Haley McNamara, Victoria Rousay, Lina Nealon, and Marcel van der Watt, *Not a Fantasy: How the Pornography Industry Exploits Image-based Sexual Abuse in Real Life*, (Washington, DC: National Center on Sexual Exploitation, forthcoming).
- ³¹ *Prague Morning*, "Czech Casting: Women Lured by Modeling Gigs, Manipulated into Shooting Porn," December 5, 2023, <https://praguemorning.cz/czech-casting-women-lured-by-modeling-gigs-manipulated-into-porn/> (accessed May 15, 2024); Samantha Cole, Emanuel Maiberg, and Anna Koslerova, "Frankenstein's Monster: Images of Sexual Abuse are Fueling Algorithmic Porn" *Vice.com*, November 10, 2020, <https://www.vice.com/en/article/akdgnp/sexual-abuse-fueling-ai-porn-deepfake-czech-casting-girls-do-porn> (accessed May 15, 2024); US Attorney's Office, Southern District of California, "Press Release: GirlsDoPorn Operator Pleads Guilty in Sex Trafficking Conspiracy," July 26, 2022, <https://www.justice.gov/usao-sdca/pr/girlsdoporn-operator-pleads-guilty-sex-trafficking-conspiracy>

(accessed May 15, 2024); Ashlee Gore and Leisha Du Preez, “GirlsDoPorn’: Online Pornography and Politics of Responsibility for Technology Facilitated Sexual Exploitation” in *The Palgrave Handbook of Gendered Violence and Technology*, ed. Anastasia Powell, Asher Flynn, and Lisa Sugiura (Cham, Switzerland: Palgrave MacMillan, 2021), <https://link.springer.com/book/10.1007/978-3-030-83734-1>.

³² John Annese, “Pornhub to Pay More than \$1.8M to Resolve GirlsDoPorn Sex Trafficking Probe: Feds,” Daily News, December 21, 2023, <https://www.nydailynews.com/2023/12/21/pornhub-to-pay-more-than-1-8m-to-resolve-girlsdoporn-sex-trafficking-probe-feds/> (accessed May 15, 2024).

³³ Lisa L. Thompson, Haley McNamara, Victoria Rousay, Lina Nealon, and Marcel van der Watt, *Not a Fantasy: How the Pornography Industry Exploits Image-based Sexual Abuse in Real Life*, (Washington, DC: National Center on Sexual Exploitation, forthcoming).

³⁴ Samantha Cole, Emmanuel Mailberg, and Anna Koslerova, “Frankenstein’s Monster:’ Images of Sexual Abuse Are Fueling Algorithmic Porn” Vice (November 10, 2020), <https://www.vice.com/en/article/akdgnp/sexual-abuse-fueling-ai-porn-deepfake-czech-casting-girls-do-porn> (accessed May 16, 2024).

³⁵ Annie Lord, “Porn Doesn’t Suffer from a ‘Piracy Problem’, It Actively Relies Upon It,” *Vice*, September 3, 2019, <https://www.vice.com/en/article/kz4899/free-youporn-redtube-pornhub-piracy-alessa-savage-ukcutegirl> (accessed April 11, 2014).

³⁶ Samantha Cole and Joseph Cox, “Inside the Underground Trade of Pirated OnlyFans Porn,” *Vice*, June 17, 2020, <https://www.vice.com/en/article/5dz3xa/onlyfans-pirated-porn-scraper-leak> (accessed April 16, 2024).

³⁷ Kayee Hanaoka, Mei Ngan, Joyce Yang, et al., *Face Analysis Technology Evaluation: Age Estimation and Verification*, (Gaithersburg, MD: National Institute of Standards and Technology, May 2024), NIST IR 8525, <https://nvlpubs.nist.gov/nistpubs/ir/2024/NIST.IR.8525.pdf> (accessed June 4, 2024).

³⁸ OpenAI.com, OpenAI and Reddit Partnership,” May 16, 2024, <https://openai.com/index/openai-and-reddit-partnership/> (accessed July 27, 2024).

³⁹ National Center on Sexual Exploitation, “Reddit: A Mainstream Contributor to Sexual Exploitation. The Dirty Dozen List 2024,” n.d., <https://endsexualexploitation.org/reddit/> (accessed July 9, 2024).

⁴⁰ Ibid.

⁴¹ United States Securities and Exchange Commission, “Amendment No. 3 to Form S-1 Registration Statement Under the Securities Act of 1933, Reddit, Inc.” March 19, 2024, <https://www.sec.gov/Archives/edgar/data/1713445/000162828024011789/reddit-sx1a3.htm> (accessed July 9, 2024).

⁴² National Center on Sexual Exploitation to Steve Huffman, Co-Founder and CEO, Reddit,” April 8, 2024, https://endsexualexploitation.org/wp-content/uploads/Reddit-Notification-Letter_DDL-2024_FINAL.pdf (accessed July 9, 2024).

⁴³ OpenAI.com, “Usage Policies,” January 10, 2024, <https://openai.com/policies/usage-policies/> (accessed May 16, 2024).

⁴⁴ Ryan Tarinelli, “Senators Rebuke Tech Execs over Child Exploitation on Social Media,” *Roll Call*, January 31, 2024, <https://rollcall.com/2024/01/31/senators-rebuke-tech-execs-over-child-exploitation-on-social-media/> (January 31, 2024).

⁴⁵ National Center on Sexual Exploitation, “Letter to Mr. Tim Cook,” March 13, 2024, https://endsexualexploitation.org/wp-content/uploads/Apple-Notification-Letter_DDL-2024_FINAL.pdf (accessed May 23, 2024).

⁴⁶ National Center on Sexual Exploitation, “Letter to Mr. Matthew Prince, Co-Founder and CEO, Cloudflare,” April 3, 2024, https://endsexualexploitation.org/wp-content/uploads/Cloudflare-Notification-Letter_DDL-2024_FINAL.pdf (accessed May 23, 2024);

⁴⁷ National Center on Sexual Exploitation, “Letter to Mr. Jason Citron, CEO and Co-Founder, Discord” April 3, 2024, https://endsexualexploitation.org/wp-content/uploads/Discord-Notification-Letter_DDL-2024_FINAL.pdf (accessed May 23, 2024).

⁴⁸ National Center on Sexual Exploitation, “Letter to Mr. Mark Zuckerberg, CEO, Meta Platforms, Inc.” April 8, 2024, https://endsexualexploitation.org/wp-content/uploads/Meta-Notification-Letter_DDL-2024_FINAL.pdf (accessed May, 23, 2024).

-
- ⁴⁹ National Center on Sexual Exploitation, “Letter to Mr. Thomas Dohmke, CEO, GitHub, and Mr. Satya Nadella, CEO, Microsoft Corporation,” April 3, 2024, https://endsexualexploitation.org/wp-content/uploads/MS-GitHub-Notification-Letter_DDL-2024_FINAL.pdf (accessed May 23, 2024).
- ⁵⁰ National Center on Sexual Exploitation, “Letter to Mr. Steve Huffman,” Co-Founder and CEO, Reddit,” April 8, 2014, https://endsexualexploitation.org/wp-content/uploads/Reddit-Notification-Letter_DDL-2024_FINAL.pdf (accessed May 23, 2024).
- ⁵¹ National Center on Sexual Exploitation, “Letter to Jack Dorsey, Chief Executive Officer and Co-Founder, and Omid Kordestani, Executive Chairman, Board of Directors, Twitter, Inc.,” https://endsexualexploitation.org/wp-content/uploads/dd_TwitterLetter_Signed_02-01-2018.pdf (accessed May 23, 2024).
- ⁵² OpenAI.com, “Usage Policies,” January 10, 2024, <https://openai.com/policies/usage-policies/> (accessed June 5, 2024).
- ⁵³ Leike, Jan, X Post, May 17, 2024, 11:57 AM, <https://x.com/janleike/status/1791498184671605209> (accessed June 6, 2024).
- ⁵⁴ Leike, Jan, X Post, May 17, 2024, 11:57 AM, <https://x.com/janleike/status/1791498182125584843> (accessed June 6, 2024).
- ⁵⁵ Leike, Jan, X Post, May 17, 2024, 11:57 AM, <https://x.com/janleike/status/1791498178346549382> (accessed June 6, 2024).
- ⁵⁶ Kevin Roose, “OpenAI Insiders Warn of a ‘Reckless’ Race for Dominance,” *The New York Times*, June 4, 2024, <https://www.nytimes.com/2024/06/04/technology/openai-culture-whistleblowers.html> (accessed June 6, 2024).
- ⁵⁷ Ibid.
- ⁵⁸ RighttoWarn.AI, “A Right to Warn about Advanced Artificial Intelligence,” June 4, 2024, <https://righttowarn.ai> (accessed June 6, 2024).
- ⁵⁹ OpenAI.com, “Usage Policies,” January 10, 2024, <https://openai.com/policies/usage-policies/> (accessed June 5, 2024).
- ⁶⁰ Angus Crawford and Tony Smith, “Illegal Trade in AI Child Sex Abuse Images Exposed,” *BBC*, June 28, 2023, <https://www.bbc.com/news/uk-65932372> (accessed June 5, 2024).
- ⁶¹ Ibid.
- ⁶² OpenAI.com, “OpenAI’s Approach to Frontier Risk. An Update for the UK AI Safety Summit,” October 26, 2023, <https://openai.com/global-affairs/our-approach-to-frontier-risk/> (accessed July 24, 2024).
- ⁶³ Stuart A. Thompson, “Uncensored Chatbots Provoke a Fracas Over Free Speech,” *The New York Times*, July 2, 2023, <https://www.nytimes.com/2023/07/02/technology/ai-chatbots-misinformation-free-speech.html> (accessed May 16, 2024).
- ⁶⁴ Tristan Harris, Twitter Post, March 10, 2023, 4:07PM, <https://x.com/tristanharris/status/1634299911872348160?s=46&t=dKA6bnISNjkb2vECGEsKw> (accessed July 23, 2024).
- ⁶⁵ Maggie Harris Deupré, “Detailed Jailbreak Gets ChatGPT to Write Wildly Explicit Smut,” *Futurism*, April 24, 2023, <https://futurism.com/jailbreak-chatgpt-explicit-smut> (accessed June 5, 2024).
- ⁶⁶ Yi Liu, Gelei Deng, Zhengzi Xu, et al., “Jailbreaking ChatGPT via Prompt Engineering: An Empirical Study,” arXiv, 2024, <https://arxiv.org/abs/2305.13860> (accessed June 6, 2024).
- ⁶⁷ Vishwam Sankaran, “ChatGPT Cooks Up Fake Sexual Harassment Scandal and Names Real Law Professor as Accused,” *Independent*, April 6, 2023, <https://www.independent.co.uk/tech/chatgpt-sexual-harassment-law-professor-b2315160.html> (accessed May 23, 2024).
- ⁶⁸ James Vincent, “OpenAI Sued for Defamation after ChatGPT Fabricates Legal Accusations against Radio Host,” *The Verge*, June 9, 2023, <https://www.theverge.com/2023/6/9/23755057/openai-chatgpt-false-information-defamation-lawsuit> (accessed June 11, 2024).
- ⁶⁹ Josh Butler, “AI Tools Could Be Used by Predators to ‘Automate Child Grooming,’ eSafety Commissioner Warns,” *The Guardian*, May 19, 2023, <https://www.theguardian.com/technology/2023/may/20/ai-tools-could-be-used-by-predators-to-automate-child-grooming-esafety-commissioner-warns> (accessed July 24, 2024).
- ⁷⁰ Santiago Lakatos, *A Revealing Picture: AI-Generated ‘Undressing’ Images Move from Niche Pornography Discussion Forums to a Scaled and Monetized Online Business*, Graphika, December 2023.

-
- ⁷¹ Santiago Lakatos, *A Revealing Picture: AI-Generated 'Undressing' Images Move from Niche Pornography Discussion Forums to a Scaled and Monetized Online Business*, Graphika, December 2023, 1.
- ⁷² Nicholas Kristof, "The Online Degradation of Women and Girls that We Meet with a Shrug," *The New York Times*, March 23, 2024, <https://www.nytimes.com/2024/03/23/opinion/deepfake-sex-videos.html?pgtype=Article&action=click&module=RelatedLinks> (accessed June 6, 2024).
- ⁷³ Ibid.
- ⁷⁴ Natasha Singer, "Teen Girls Confront an Epidemic of Deepfake Nudes in Schools," *The New York Times*, April 8, 2024, <https://www.nytimes.com/2024/04/08/technology/deepfake-ai-nudes-westfield-high-school.html> (accessed June 6, 2024).
- ⁷⁵ Alexandra S. Levine, "'I Want that Sweet Baby': AI-Generated Kids Draw Predators on TikTok and Instagram," *Forbes*, May 20, 2024, <https://www.forbes.com/sites/alexandrlevine/2024/05/20/ai-generated-kids-tiktok-instagram-social-media-child-safety-predators/?sh=fcb0b6726cb7> (accessed June 5, 2024).
- ⁷⁶ Ibid.
- ⁷⁷ Internet Watch Foundation, *How AI is being Abused to Create Child Sexual Abuse Imagery*, (Cambridge, UK: Internet Watch Foundation, October 2023), <https://www.iwf.org.uk/media/q4zll2ya/iwf-ai-csam-report-public-oct23v1.pdf> (accessed June 5, 2024).
- ⁷⁸ Ibid.
- ⁷⁹ Internet Watch Foundation, *What Has Changed in the AI CSAM Landscape? AI CSAM Report Update in Conjunction with our Oct 23 Report*, July 2024, <https://www.iwf.org.uk/media/drufozvi/iwf-ai-csam-report-update-public-jul24v12.pdf> (accessed July 23, 2024).
- ⁸⁰ Angus Crawford and Tony Smith, "Illegal Trade in AI Child Sex Abuse Images Exposed," *BBC*, June 28, 2023, <https://www.bbc.com/news/uk-65932372> (accessed June 5, 2024).
- ⁸¹ Alexandra S. Levine, "AI Is Creating an Influx of Child Sex Abuse Images, Data Shows," *Forbes*, April 15, 2024, https://www.forbes.com/sites/alexandrlevine/2024/04/15/ai-is-creating-an-influx-of-child-sex-abuse-images-ncmec-data/?wpisrc=nl_technology202&sh=cf10ce57ad3a (accessed June 5, 2024).
- ⁸² Childlight – Global Safety Institute, *Into the Light Index on Child Sexual Exploitation and Abuse Globally: 2024 Report*, (Edinburgh: Childlight, 2024), <https://childlight.org/sites/default/files/2024-05/into-the-light.pdf> (accessed June 6, 2024).
- ⁸³ Elizabeth J. Letourneau, William E. Eaton, Judith Bass, "The Need for a Comprehensive Public Health Approach to Preventing Child Sexual Abuse," *Public Health Reports* 129, no. 3 (2014): 222–228, <https://doi.org/10.1177/003335491412900303>.
- ⁸⁴ Michael H. Keller and Gabriel J.X. Dance, "The Internet Is Overrun with Images of Child Sexual Abuse. What Went Wrong?" *The New York Times*, September 29, 2019, <https://www.nytimes.com/interactive/2019/09/28/us/child-sex-abuse.html> (accessed June 25, 2024).
- ⁸⁵ Michael Barbaro et al., interview with Michael H. Keller and Gabriel J.X. Dance, "A Criminal Underworld of Child Abuse, Part 1," February 19, 2020, in *The Daily* podcast, produced by *The New York Times*, <https://www.nytimes.com/2020/02/19/podcasts/the-daily/child-sex-abuse.html> (accessed June 25, 2024).
- ⁸⁶ Kathryn Seigfried-Spellar and Marcus K. Rogers, "Does Deviant Pornography Use Follow a Guttman-Like Progression?" *Computers in Human Behavior* 29, no. 5 (2013): 1997–2003, <https://doi.org/10.1016/j.chb.2013.04.018>.
- ⁸⁷ Kathryn Seigfried-Spellar, "Deviant Pornography Use: The Role of Early-Onset Adult Pornography Use and Individual Differences," *International Journal of Cyber Behavior, Psychology and Learning* 6, no. 3 (2016): 34–47, <https://doi.org/10.4018/IJCBPL.2016070103>.
- ⁸⁸ Francis Fortin, Sarah Paquette, and Benoit Dupont, "From Online to Offline Offending: Episodes and Obstacles," *Aggression and Violent Behavior* 39 (2018): 33–41, <https://doi.org/10.1016/j.avb.2018.01.003>; Francis Fortin and Jean Proulx, "Sexual Interest of Child Sexual Exploitation Material (CSEM) Consumers: Four Patterns of Severity over Time," *International Journal of Offender Therapy and Comparative Criminology* 63, no. 1 (2019): <https://doi.org/10.1177/0306624X18794135>; Hannah L. Merdian et al., "'So Why did You do it?': Explanations Provided by Child Pornography Offenders," *Sexual Offender Treatment* 8, no. 1 (2013): 1–19, <https://hdl.handle.net/10289/7976>; Michael C. Seto and Angela W. Eke, "Correlates of Admitted Sexual Interest in Children among Individuals Convicted of Child Pornography Offenses," *Law and Human Behavior* 41, no. 3 (2017): 305–313, <https://doi.org/10.1037/lhb0000240>;

-
- Michael C. Seto, Lesley Reeves, and Sandy Jung, "Explanations Given by Child Pornography Offenders for Their Crimes," *Journal of Sexual Aggression* 16, no. 2 (2010):169–180, <https://doi.org/10.1080/13552600903572396>; Ethel Quayle and Max Taylor, "Child Pornography and the Internet: Perpetuating a Cyle of Abuse," *Deviant Behavior* 23, no 4. (2002): 331–361, <http://dx.doi.org/10.1080/01639620290086413>.
- ⁸⁹ Jennifer A. McCarthy, "Internet Sexual Activity: A Comparison between Contact and Non-Contact Child Pornography Offenders," *Journal of Sexual Aggression* 16, no. 2 (2010): 181-195, doi:10.1080/13552601003760006.
- ⁹⁰ Francis Fortin and Jean Proulx, "Sexual Interest of Child Sexual Exploitation Material (CSEM) Consumers: Four Patterns of Severity Over Time," *International Journal of Offender Therapy and Comparative Criminology* 63, no. 1 (2019): 55–76, <https://doi.org/10.1177/0306624X18794135>.
- ⁹¹ Alien Wéry and Joel Billieux, "Online Sexual Activities: An Exploratory Study of Problematic and Non-problematic Usage Patterns in a Sample of Men," *Computers in Human Behavior* 56 (2016): 257–266, <https://doi.org/10.1016.chb.2015.11.046>.
- ⁹² Simone Kühn and Jürgen Gallinat, "Brain Structure and Functional Connectivity Associated with Pornography Consumption: The Brain on Porn," *JAMA Psychiatry* 71, no. 7 (2014): 827–834, <https://doi.org/10.1001/jamapsychiatry.2014.93>.
- ⁹³ Web ranking and site visit data collated from Similar Web results as of January 2024.
- ⁹⁴ Michael L. Bourke and Andres E. Hernandez, "The 'Butner Study' Redux: A Report on the Incidence of Hands-On Child Victimization by Child Pornography Offenders," *Journal of Family Violence* 24 (2009): 183–191, <https://doi.org/10.1007/s10896-008-9219-y>; Michael L. Bourke, Letter to Judge Patti B. Saris, Chair, US Sentencing Commission, May 17, 2012, https://www.ussc.gov/sites/default/files/pdf/amendment-process/public-hearings-and-meetings/20120215/Testimony_15_Bourke.pdf (accessed June 25, 2024).
- ⁹⁵ World Health Organization, "Global Fact Sheet, Violence against Women Prevalence Estimates, 2018," (2021), <https://www.who.int/publications/i/item/WHO-SRH-21.6> (accessed June 13, 2024).
- ⁹⁶ World Health Organization, "Sexual Violence," (2022), <https://apps.who.int/violence-info/sexual-violence/> (accessed June 13, 2024).
- ⁹⁷ World Health Organization, Department of Sexual and Reproductive Health and Research, "Violence Studies," (2022), <https://apps.who.int/violence-info/studies?area=sexual-violence&aspect=consequences&consequence-level=Social%20and%20behavioural%20problems> (accessed June 13, 2024).
- ⁹⁸ Malachi Willis, Ana J. Bridges, and Chyng Sun, "Pornography Use, Gender, and Sexual Objectification: A Multinational Study," *Sexuality & Culture* 26 (2022): 1298–1313, <https://doi.org/10.1007/s12119-022-09943-z>; Jochen Peter and Patti M. Valkenburg, "Adolescent's Exposure to Sexually Explicit Internet Material and Notions of Women as Sex Objects: Assessing Causality and Underlying Processes," *Journal of Communication* 59 (2009): 407–433, <https://doi.org/10.1111/j.1460-2466.2009.01422.x>.
- ⁹⁹ Tomi-Ann Roberts, Rachel M. Calogero, and Sarah J. Gervais, "Objectification Theory: Continuing Contributions to Feminist Psychology," in *APA Handbook on the Psychology of Women Vol. 1 History, Theory, and Battlegrounds*, eds. in chief, Cheryl B. Travis and Jacquelyn W. White (Washington, DC: American Psychological Association, 2018), 256.
- ¹⁰⁰ Bhuvanesh Awasthi, "From Attire to Assault: Clothing, Objectification, and De-humanization – A Possible Prelude to Sexual Violence?" *Frontiers in Psychology* 8 (2017), <https://doi.org/10.3389/fpsyg.2017.00338>.
- ¹⁰¹ Niki Fritz, Vinny Malic, Bryant Paul, et al., "A Descriptive Analysis of the Types, Targets, and Relative Frequency of Aggression in Mainstream Pornography," *Archives of Sexual Behavior* 49 (2020): 3041–3053, <https://doi.org/10.1007/s10508-020-01773-0>.
- ¹⁰² Fiona Vera-Gray et al., "Sexual Violence as a Sexual Script in Mainstream Online Pornography," *The British Journal of Criminology* (2021): 1-18, <https://doi.org/10.1093/bjc/azab035>.
- ¹⁰³ Michael B. Robb and Supreet Mann, *Teens and Pornography*, Common Sense Media (2022), <https://www.common Sense Media.org/sites/default/files/research/report/2022-teens-and-pornography-final-web.pdf> (accessed June 25, 2024).
- ¹⁰⁴ Michele L. Ybarra et al., "X-Rated Material and Perpetration of Sexually Aggressive Behavior among Children and Adolescents: Is There a Link?" *Aggressive Behavior* 37, no. 1 (2011): 1–18, <https://doi.org/10.1002/ab.20367>.

-
- ¹⁰⁵ Paul J. Wright, Robert S. Tokunaga, and Ashley Kraus, "A Meta-Analysis of Pornography Consumption and Actual Acts of Sexual Aggression in General Population Studies," *Journal of Communication* 66, no. 1 (2016): 183–205, <https://doi.org/10.1111/jcom.12201>.
- ¹⁰⁶ Michael C. Seto and Martin L. Lalumiere, "What is So Special about Male Adolescent Sexual Offending? A Review and Test of Explanations through Meta-Analysis," *Psychological Bulletin* 136, no. 4 (2010): 526–575, <https://doi.org/10.1037/a0019700>; Michele L. Ybarra and Kimberly Mitchell, "Prevalence Rates of Male and Female Sexual Violence Perpetrators in a National Sample of Adolescents," *JAMA Pediatrics* 167 no. 12 (2013): 1125–1134, [doi:10.1001/jamapediatrics.2013.2629](https://doi.org/10.1001/jamapediatrics.2013.2629). [**Note: sample age ranges were from 14- to 21-years-old; impact was attributable to exposure to violent pornography**].
- ¹⁰⁷ Michele L. Ybarra and Richard E. Thompson, "Predicting the Emergence of Sexual Violence in Adolescence," *Prevention Science* (2017), <https://doi.org/10.1007/s11121-017-0810-4>.
- ¹⁰⁸ Michele L. Ybarra and Kimberly Mitchell, "Exposure to Internet Pornography among Children and Adolescents: A National Survey," *CyberPsychology & Behavior* 8, no. 5 (2005): 473–486, <https://doi.org/10.1089/cpb.2005.8.473> (Note: physical abuse included being hit, beaten, kicked, physically abused by any grown-up taking care of them, and sexual victimization was defined as being forced or made to do sexual things by someone, including someone they didn't know or even someone they knew well); Fanghong Dong, Fenglin Cao, Peixia Cheng, et al., "Prevalence and Associated Factors of Poly-Victimization in Chinese Adolescents," *Scandinavian Journal of Psychology* 54, no. 5 (2013): 415–422, <https://onlinelibrary.wiley.com/doi/abs/10.1111/sjop.12059>.
- ¹⁰⁹ Jessica Laird, Bianca Klettle, Kate Hall, et al., "Demographic and Psychological Factors Associated with Child Sexual Exploitation. A Systematic Review and Meta-analysis," *JAMA Network Open* 3, no. 9 (2020):1–17,
- ¹¹⁰ World Health Organization, "HIV and AIDS," July 13, 2023, <https://www.who.int/news-room/fact-sheets/detail/hiv-aids> (accessed June 20, 2024).
- ¹¹¹ Ibid.
- ¹¹² John Stover and Yu Teng, "The Impact of Condom Use on the HIV Epidemic," [version 2; peer review: 2 approved], *Gates Open Research* 5:91 (2022), <https://doi.org/10.12688/gatesopenres.13278.2>.
- ¹¹³ Paul J. Wright, "Pornography Consumption and Condomless Sex among Emerging U.S. Adults: Results from Six Nationally Representative Surveys," *Health Communication* 37, no. 14 (2022): 1740–1747, <https://doi.org/10.1080/10410236.2021.1917745>.
- ¹¹⁴ Corita R. Grudzen, Gery Ryan, William Margold, et al., "Pathway to Health Risk Exposure in Adult Film Performers," *Journal of Urban Health: Bulletin of the New York Academy of Medicine* 86, no. 1 (2008): 73, <https://doi.org/10.1007/s11524-008-9309-4>.
- ¹¹⁵ Paul J. Wright, Debby Herbenick, and Bryant Paul, "Casual Condomless Sex, Range of Pornography Exposure, and Perceived Pornography Realism," *Communication Research* 49, no. 4 (2022): 588, <https://doi.org/10.1177/00936502211003765>.
- ¹¹⁶ Robert S. Tokunaga, Paul J. Wright, and Laurens Vangeel, "Is Pornography Consumption a Risk Factor for Condomless Sex?" *Human Communication Research* 46 (2020): 273–299, <https://doi.org/10.1093/hcr/hqaa005>.
- ¹¹⁷ Paul J. Wright, "Pornography Consumption and Condomless Sex among Emerging U.S. Adults: Results from Six Nationally Representative Surveys," *Health Communication* 37, no. 14 (2022): 1740–1747, <https://doi.org/10.1080/10410236.2021.1917745>; Emily F. Galper, Laura Widman, Julia Brasileiro, et al., "Adolescents' Pornography Viewing Frequency and Its Relationship with Condom Attitudes," *Sexual Health* 21 (2024), <https://doi.org/10.1071/SH24025>.
- ¹¹⁸ Paul J. Wright, Debby Herbenick, and Bryant Paul, "Casual Condomless Sex, Range of Pornography Exposure, and Perceived Pornography Realism," *Communication Research* 49, no. 4 (2022): 588, <https://doi.org/10.1177/00936502211003765>.
- ¹¹⁹ M. Reinert, D. Fritze, and T. Nguyen, *The State of Mental Health in America* (Alexandria, VA: Mental Health America, October 2022), <https://mhanational.org/sites/default/files/2023-State-of-Mental-Health-in-America-Report.pdf?eType=ActivityDefinitionInstance&eId=5768b343-b128-4de9-a180-20ed43f570d4> (accessed June 25, 2024).
- ¹²⁰ Ibid.
- ¹²¹ Zara Abrams, "Kid's Mental Health Is in Crisis. Here's What Psychologists are Doing to Help," American Psychological Association, January 1, 2023, <https://www.apa.org/monitor/2023/01/trends-improving-youth-mental-health> (accessed June 25, 2024).

-
- ¹²² Megan S. C. Lim, Paul A. Agius, Elise R. Carroette, et al., "Young Australians Use of Pornography and Associations with Sexual Risk Behaviours," *Australian and New Zealand Journal of Public Health* 41, no. 4 (2017): 438–443, <https://doi.org/10.1111/1753-6405.12678>.
- ¹²³ Brian J. Willoughby, Jason S. Carroll, Larry J. Nelson, et al., "Associations between Relational Sexual Behaviour, Pornography Use, and Pornography Acceptance among US College Students," *Culture, Health & Sexuality* 16, no. 9 (2014): 1052–69, <https://doi.org/10.1080/13691058.2014.927075>.
- ¹²⁴ Manuel Mennig, Sophia Tennie, and Antonia Barke, "Self-Perceived Problematic Use of Online Pornography is Linked to Clinically Relevant Levels of Psychological Distress and Psychopathological Symptoms," *Archives of Sexual Behavior* 51 (2022): 1313–1321, <https://doi.org/10.1007/s10508-021-02101-w>.
- ¹²⁵ Manuel Mennig, Sophia Tennie, and Antonia Barke, "Self-Perceived Problematic Use of Online Pornography is Linked to Clinically Relevant Levels of Psychological Distress and Psychopathological Symptoms," *Archives of Sexual Behavior* 51 (2022): 1319, <https://doi.org/10.1007/s10508-021-02101-w>.
- ¹²⁶ Michele L. Ybarra and Kimberly Mitchell, "Exposure to Internet Pornography among Children and Adolescents: A National Survey," *CyberPsychology & Behavior* 8, no. 5 (2005): 473–486, <https://doi.org/10.1089/cpb.2005.8.473>.
- ¹²⁷ Simone Kühn and Jürgen Gallinat, "Brain Structure and Functional Connectivity Associated with Pornography Consumption," *JAMA Psychiatry* 71, no. 7 (2014): 827–834, <https://doi.org/10.1001/jamapsychiatry.2014.93>.
- ¹²⁸ Valerie Voon, Paula Banca, Tatyana R. Lapa, et al., "Neural Correlates of Sexual Cue Reactivity in Individuals with and without Compulsive Sexual Behaviors," *PLOS ONE* 9, no. 7 (2014): 1–10, <https://doi.org/10.1371/journal.pone.0102419>.
- ¹²⁹ Mateusz Gola, Małgorzata Wordecha, Guillaume Sescousse, Micha Lew-Starowicz, Bartosz Kossowski, Marek Wypych, Scott Makeig, Marc N. Potenza, and Artur Marchewka, "Can Pornography be Addictive? an fMRI Study of Men Seeking Treatment for Problematic Pornography Use," *Neuropsychopharmacology* 42, no. 10 (2017): 2021–2031, <https://doi.org/10.1038/npp.2017.78>.
- ¹³⁰ Paris Adult Theatre I v. Slaton, 413 U.S. 49 (1973), <https://supreme.justia.com/cases/federal/us/413/49/>.