

DRAWING **THE LINE**

About Drawing the Line

Drawing the Line Between Personal Expression and Lived Abuse is an ongoing program of research and advocacy work intended to clearly delineate the boundary between personal expression and lived abuse in the context of digital content regulation and moderation.

It is grounded in a set of guiding principles, the Drawing the Line Principles, which are intended to inform the development of rights-respecting approaches to sexual expression, censorship, and content moderation, primarily focusing on digital spaces. They aim to support advocates, policymakers, researchers, and platform operators in addressing the harms of overbroad censorship and discrimination, while meaningfully distinguishing consensual sexual expression from abuse.

The Drawing the Line Principles draw from a range of intersecting normative and legal frameworks, including the international human rights system—particularly rights to freedom of expression, privacy, equality and non-discrimination, and the highest attainable standard of health—as well as feminist, queer, sex worker rights, and digital justice movements. They are informed by civil society calls for platform accountability, transparency, and user empowerment, and by ongoing efforts to decolonize and democratize digital governance. This paper also engages with harm reduction principles, public health perspectives, and community-centered knowledge on safety, pleasure, and consent.

The principles have been developed collaboratively by a diverse group of contributors from domains such as online trust and safety, the arts, human rights and criminal justice, gender and sexuality studies, and linguistics, and are currently open for broader endorsement.

Context and Urgency

Digital spaces have become essential arenas for sexual exploration, identity formation, education, community, and advocacy. Yet increasingly, these same spaces are subject to intensified surveillance, censorship, and policing of sexual content, in its broadest sense. Legislative proposals, regulatory actions, and platform policies—often in the name of protecting children, combating trafficking, or maintaining "decency"—are routinely used to suppress lawful and consensual sexual expression. These measures disproportionately impact LGBTQ+ communities, sex workers, sex educators, abuse survivors, racial and ethnic minorities, and others whose sexual speech or existence is already stigmatized. In parallel, genuine

abuse—including image-based sexual violence, harassment, and exploitation—continues to thrive in many online environments. The need for targeted, rights-based solutions has never been more urgent.

The Drawing the Line Principles seek to reframe dominant narratives that pathologize sexual speech and to promote regulatory and governance models that center consent, autonomy, and inclusion. While it acknowledges the existence of harmful, coercive, or non-consensual conduct involving sex, the principles here are explicitly concerned with defending consensual expression from overbroad censorship. The principles do not aim to provide a comprehensive framework for addressing online sexual harms—an important but distinct project—but rather to assert the rights and realities of sexual expression in public and semi-public digital space. Where reference is made to abuse, violence, or exploitation, this is to draw contrast with lawful expression and to highlight the necessity of clear distinctions in law and policy.

Scope

This paper focuses on sexual expression in digital environments: that is, the consensual articulation or depiction of sex, sexuality, or desire, whether through text, imagery, video, or symbolic forms. The scope includes user-generated content on social media platforms and websites, digitally mediated sexual communication (e.g., sexting), sex worker content and advertising, sexual health information, erotic art and performance, and expressions of gender and sexual identity.

Creative expression and experience-sharing in digital spaces differ from in-person contexts because online content is more visible, permanent, and easily misinterpreted out of context. The moderation and regulation of digital spaces is not easy. A great deal of content moderation is done through automation, either by simple lexical triggers or more sophisticated AI systems that may not be properly trained. Humans in content moderation are often poorly trained and paid, and come to the work with their own biases. This heightened scrutiny makes digital expression especially vulnerable—particularly for marginalized communities—underscoring the need for clear distinctions in policy.

While it is possible to do this work well, it does require effort and expertise in automation, as well as human-in-loop, especially including survivors in the process of creating policy and evaluating the effectiveness and fairness of automation. Different digital spaces may have different standards of behavior based on their purpose and users, yet we believe that it is possible to work within those ideas to foster creativity and personal sharing while avoiding over-aggressive censorship.

Principle 1: Expression is Not Abuse

Fictional work and creative expression should not be conflated with acts of sexual abuse against others in policy, enforcement, or language. Collapsing this distinction risks diluting the seriousness of real abuse and undermining the rights of creators and

communities alike, while also funneling resources and attention away from meaningful efforts to prevent and combat abuse.

This principle challenges the growing tendency to treat fictional, metaphorical, or symbolic forms of expression as tantamount to actual harm. While some content may be offensive, disturbing, or controversial, the threshold for considering it abusive must remain firmly grounded in evidence of real-world harm. Mischaracterizing expression as abuse undermines both freedom of speech and the integrity of abuse prevention efforts.

Sexual expression is a core component of human dignity and identity. The ability to explore, communicate, and share sexual expression—within the bounds of legality and consent—is fundamental to the exercise of personal autonomy and freedom of expression. Digital platforms have become an indispensable space for sexual speech and community-building, especially for those whose desires or identities are stigmatized or marginalized in offline contexts.

In contrast, sexual abuse refers to acts of a sexual nature committed against a person without their consent or capacity to consent, involving coercion, manipulation, or exploitation. It includes but is not limited to child sexual abuse, sexual assault, and non-consensual creation or distribution of sexual imagery involving real individuals. Sexual abuse causes real harm and trauma, and must be distinguished from fictional, artistic, or consensual representations of sexuality.

Yet around the world, individuals and communities—particularly those already marginalized—are routinely treated as dangerous simply for expressing their sexuality online. Consensual sexual speech, imagery, and identity are often swept up in policies designed to address genuine harm, resulting in the over-removal of lawful content and the policing of sexual expression under the guise of safety.

This conflation of sexual expression with abuse stems from moralistic and protectionist frameworks that prioritize perceived decency or risk-aversion over autonomy, consent, and evidence. Efforts to prevent sexual abuse—particularly involving minors—are crucial and widely supported. But when protection is pursued through sweeping or vague restrictions on lawful expression, it not only fails to prevent actual harm but also endangers fundamental rights. Such approaches fail to distinguish between consensual expression and exploitative behavior, leading to the suppression of entire communities, including LGBTQ+ people, sex workers, sexual health educators, and others whose identities or work involve open engagement with sex and sexuality.

This principle calls out three ways in which expression can be conflated with abuse: in policy, enforcement, and language:

- In policy, laws and regulations may define sexual abuse too broadly or ambiguously, sweeping in fictional, artistic, or educational content that involves no real victim. Consistent with the principle that expression is not abuse, a better practice for policymakers is to clearly define sexual abuse as involving non-consensual acts

involving real individuals, and to explicitly exclude fictional, simulated, or artistic content from abuse classifications unless it involves actual harm.

- In enforcement, authorities and platforms may interpret or apply those policies without nuance—blocking, banning, or prosecuting expressive content under the guise of abuse prevention. Better practices for enforcement teams and agencies include avoiding the use of low-evidence thresholds (like algorithmic flags or decontextualized imagery) as grounds for punitive action, building in transparent appeals processes and due process protections for creators and users whose content is challenged, and prioritizing resources for identifying and addressing actual cases of abuse involving real individuals.
- And in language, the use of emotionally charged or misleading terms (like “child sexual abuse material” or CSAM to describe fictional works) distorts public understanding, undermines support for actual survivors, and fuels stigma against creators and communities. Terms like AI-generated CSAM have already become ubiquitous (eg. Safe Online, 2025, p. 13; ECPAT, 2025; Thorn & All Tech is Human, 2024), but this usage should be resisted except in narrowly defined cases where real individuals are directly harmed—for instance, deepfake imagery produced by nudify apps that targets real people. Accurate, context-sensitive language is essential to uphold the integrity of anti-abuse efforts and protect lawful expression.

To protect sexual rights in digital spaces, we must draw clear and principled distinctions between harm and expression. Lawful sexual content that is consensually created and shared should not be treated as inherently harmful. Platforms, governments, and civil society must reject narratives that pathologize sexual expression or equate it with exploitation. Instead, they must center consent, uphold the right to freedom of expression, and implement safeguards that effectively target abuse while preserving space for lawful, consensual sexual speech.

Principle 2: Defend Creative and Cultural Expression

Fiction, art, fantasy, personal narratives, and other creative content are essential components of human expression, cultural exploration and critique, and personal growth. Content reflecting lived experiences and creative expression which includes queerness, kink, LGBTQ identities, survivor identity, and racial or cultural identity must not be restricted or stigmatized based on those factors.

Creative expression has long served as a conduit for exploring and communicating complex aspects of the human experience. It encompasses not only entertainment and art but also political discourse, identity formation, and cultural memory. Protecting the right to create and disseminate expressive works—especially those that deal with sensitive, controversial, or marginalized perspectives—is central to the preservation of democratic values and the advancement of human rights.

International human rights law recognizes the freedom of expression as a cornerstone of democratic society. Article 19 of the International Covenant on Civil and Political Rights (ICCPR)

affirms the right to "seek, receive and impart information and ideas of all kinds... regardless of frontiers." The UN Human Rights Committee's General Comment No. 34 on Article 19 explicitly states that this includes "expressive content in the form of art," and that the right to freedom of expression encompasses expression that may be "deeply offensive" to some (CCPR/C/GC/34, 2011).

Literature, visual art, film, music, and interactive media have long been central to struggles for civil rights, queer visibility, sexual autonomy, and survivor justice. These works can offer catharsis, foster empathy, challenge dominant ideologies, and reimagine social roles. But they are also vulnerable to moral panic and censorship. From book bans targeting queer youth literature to the Comics Code Authority and restrictions on hip-hop lyrics in the 1990s, history shows how fears about public morality or child protection are often used to suppress valuable cultural expression. Today, those patterns continue, especially online.

In particular, expression related to queerness, kink, survivorhood, or minority cultural experiences is often disproportionately stigmatized or censored under the guise of protecting public morality or safety. Such restrictions frequently reflect underlying societal prejudices rather than objective harm assessments. Scholars have documented how such content is subject to disproportionate scrutiny, especially on digital platforms that operate transnationally and are influenced by varying regional norms (GLAAD, 2025; Noble, 2018). Examples abound:

- Queer expression has historically been suppressed through obscenity laws and continues to face platform-level restrictions under vague or inconsistent content policies. A series of reports from GLAAD published since 2021 have found that LGBTQ+ creators are frequently demonetized or had their content restricted even when it did not violate community guidelines (2025).
- Kink and BDSM-related content, even when educational or entirely fictional, is often flagged or removed by automated systems or subjected to shadow bans. This occurs despite such content being legal to enact in person, and integral to the sexual identities and communities it represents. This April 2025 in Australia for example, a female novelist was arrested on child abuse charges over an 18+ erotic novel that she wrote (Beazley, 2025). Over the last decade in countries including Canada (Carter, 2020), France (franceinfo & AFP, 2025), and the United Kingdom (Malcolm, 2019), other authors have faced child abuse charges over novels and comic books.
- Across the world from Costa Rica to Russia, LGBTQ+ people and children are disproportionately represented among those who are arrested for victimless obscenity crimes (Gazeta, 2019; Sandí, 2019).
- Survivors are not exempt from this either, as survivor narratives—particularly those that include references to trauma, abuse, or recovery—can be misinterpreted as glorifying or triggering, leading to restrictions that silence survivors rather than supporting them. For example, one graphic novel targeted was a survivor's memoir of incestual abuse (Af viceland, 2006).

- Librarians and archivists, including some who are also survivors (Ward, 2008) are also facing charges over virtual child abuse crimes (Hixenbaugh et al., 2024). Some have received punishments higher than those received by convicted sex traffickers and molesters (SHG, 2021; Swan, 2023).
- In April 2025 in the United States a game publisher was forced to withdraw a fantasy erotic game, while in other countries the game was officially censored (Wilde, 2025). By July 2025, this had widened to include a broader censorship sweep against games with sexual themes, covering multiple online platforms including Steam and itch.io (Farokhmanesh, 2025).
- However, these standards are applied unevenly. Big budget works from mainstream studios such as *Game of Thrones* and *Euphoria* win acclaim and awards for depicting scenarios that would create censure and legal consequences for creators and fans from less privileged groups.

Internationally, countries are being pressured to add new criminal offences for virtual sex offences (Lanzarote Committee, 2024; UN Committee on the Rights of the Child, 2019). Some jurisdictions, such as Australia, do not even record the difference between virtual and real image-based offences in official statistics (Malcolm, 2024).

This lack of distinction at the legal level often sets the stage for similarly blunt instruments in enforcement and content moderation. For example, recent research demonstrates that the conflation of real Child Sexual Abuse Materials with animation leads to unsuccessful moderation practices on adult platforms that might unfairly target uploaders and content creators (Petit, 2025).

These trends echo long-standing debates in feminist and legal scholarship. While some critiques of pornography or sexualized media have sought to challenge exploitation, others have reinforced systems of censorship that silence sexual speech and reassert patriarchal control (MacKinnon, 1989; Strossen, 2000). As the digital public square becomes the primary space for discourse and art, ensuring that content policies do not impose disproportionate burdens on artists, minorities, or survivors becomes both a freedom of expression issue and a matter of social equity.

Principle 3: Respect Survivor Autonomy

Survivors of sexual violence deserve respect, support, and autonomy. The way that content thresholds are set in digital spaces can stigmatize and disenfranchise survivors, or trivialize their lived experience. Ethical content policies should prioritize the agency of survivors and not co-opt them into supporting overbroad censorship measures.

Efforts to improve online safety must center the rights and dignity of survivors—not just as subjects of concern, but as rights-holders with agency and insight. Too often, survivors are invoked as a justification for restrictive policies without being consulted about what safety

actually means to them. Measures framed as protective can easily veer into paternalism, especially when they silence or delegitimize survivor voices.

Content moderation policies, for instance, may suppress personal narratives, educational resources, or community support spaces that are vital to healing and empowerment. It is known that survivors are overrepresented both as consumers and as creators of works that deal with dark sexual themes (Rouse et al., 2023). Algorithms that automatically flag sexual content rarely distinguish between exploitative material and survivor-led expression or advocacy. This erasure not only retraumatizes individuals but undermines collective efforts to break silence around abuse.

Respecting survivor autonomy requires more than avoiding harm; it demands active participation and co-creation. Survivors should be involved in shaping safety policies, content standards, and enforcement practices—especially those that claim to act in their name. Their experiences should inform how platforms define thresholds for harmful content, build trust and safety tools, and evaluate impact.

A consent-based, survivor-led approach recognizes that safety is not one-size-fits-all. It affirms the right of survivors to speak, share, organize, and heal on their own terms—rather than being made invisible by rules that prioritize institutional liability over lived experience.

While not expressly referenced in this principle, it is also essential to address the misallocation of resources in responses to sexual harm. Public and private actors increasingly devote funding and enforcement capacity to the suppression of fictional content under the banner of safety, while real-world support systems for survivors remain grossly underfunded. Fewer than 4% of real sexual assaults ever result in a felony conviction (Walinchus et al., 2025), and only 3.5% of CSAM reports are ever even investigated by legal authorities (Bischoff, 2021). Of equal or greater significance, only miniscule funding is made available for resources for the prevention of sexual abuse (Letourneau, 2022), and practitioners of sexual abuse prevention are subject to stigmatizing attacks (Walker, 2023).

Respecting survivor autonomy means recognizing survivors not as symbols or political tools, but as full participants in public discourse. Platforms, policymakers, and civil society must invite survivors into decision-making, avoid using them as rhetorical shields for overreach, and ensure that expressions of healing and advocacy are protected rather than penalized. True support requires investment in prevention, accountability, and survivor-led spaces—not policies that erase their voices under the guise of protection.

Principle 4: Support Evidence-Based Safety and Prevention Practices

Effective prevention of sexual violence and abuse requires investments in public health, education, consent culture, economic justice, and social services. The censorship of legal content cannot be justified as an evidence-based abuse prevention measure.

Meaningful progress in protecting people and building a safer society will be made through anchoring public discourse and policy in facts, not fear.

Censorship and surveillance, even when well-intentioned, are often reactive tools that ignore the root causes of harm. Sustainable safety requires confronting the structural and social conditions that allow abuse to occur and persist—such as inequality, stigma, and a lack of comprehensive sex education or mental health care. Misguided efforts to police lawful expression may create a false sense of action, while diverting attention from the systems that fail survivors and communities alike.

Evidence consistently shows that community safety improves through trauma-informed, intersectional approaches that reduce both victimization and perpetration. These include access to sex-positive, inclusive education; culturally competent mental health support; economic empowerment; and care models that affirm consent, dignity, and agency.

Policy shaped by fear or moral panic often lacks such grounding. It is typically driven by unsubstantiated claims of harm or by selectively citing research to emphasize hypothetical risks while ignoring considerations of literary, artistic, political, or scientific value, resulting in policy shaped more by moral panic than by evidence (Lievesley et al., 2023, p. 401). When safety interventions rely on selectively cited or misrepresented research—while ignoring the broader social science consensus—they risk harming the very communities they purport to protect. LGBTQ+ youth, sex workers, and communities of color are disproportionately targeted by surveillance-heavy or punitive safety models that have little evidentiary support and often reproduce existing social injustices.

Representations of harm—such as fictional media dealing with trauma, abuse, or sexuality—can indeed have negative effects, such as reinforcing stereotypes or retraumatizing individuals. But these cultural harms are fundamentally different from direct acts of abuse, and they require fundamentally different tools: education, critical media literacy, and open debate, not criminalization.

To truly advance digital safety, we must hold interventions to the same standards we demand of any public policy: transparency, peer-reviewed research, and rigorous evaluation. An evidence-based approach focuses on measurable outcomes, values interdisciplinary collaboration, and remains accountable to the communities most affected. It resists the pull of moral panic and focuses instead on what works to create safer, more equitable digital spaces for all.

Principle 5: Create Robust Digital Due Process Practices

Rules restricting lawful creative content must be clearly articulated and interpretable and meet a high bar of justification and proportionality. Creators, researchers, and users must be protected from intrusive, disingenuous, or coercive demands for their private

data in the course of content moderation. Platforms and regulators must provide notice of content removal decisions and the opportunity to challenge them.

This principle states the need for both substantive and procedural safeguards against the over-censorship of sexual expression. First, substantive rules restricting lawful content must be clearly defined, consistently applied, and subject to meaningful oversight. Blanket or vague restrictions—particularly those premised on indirect or speculative harm—fail to meet international human rights standards. Platforms and regulators should limit content restrictions to material shown to cause direct harm, especially to children, and ensure that measures are evidence-based and proportionate to the harm addressed.

Another key element of substantive protection is the recognition of context. When moderating lawful sexual content, platforms must account for the circumstances of its creation and publication. Whether the material is fictional, educational, artistic, therapeutic, or produced and shared with informed consent should meaningfully influence moderation decisions. Factors such as consent, intent, and the intended audience are essential in distinguishing between content that is potentially harmful and that which serves legitimate expressive, informative, or healing purposes.

Second, procedural safeguards are essential. Users must be notified when content is removed or accounts are penalized, and provided with a clear explanation of the alleged violation, the evidence involved, and an accessible, timely, and fair appeal process. Where significant penalties are involved—such as account suspension or listing in shared industry databases—decisions should be subject to human review and mechanisms for external oversight.

Surveillance and data collection practices must also meet a high threshold of necessity and proportionality. Content creators and users should not be subjected to intrusive, coercive, or unnecessary demands for personal data, such as identity documents or biometric scans, particularly when the content in question is lawful. Systems for age assurance, content filtering, and trust scoring should be transparently disclosed, independently evaluated, and demonstrably effective at reducing harm without producing discriminatory or chilling effects.

Remedies must be effective and accessible. Survivors of image-based abuse should be prioritized in takedown systems. At the same time, platforms should empower users with tools to filter sexual content they do not wish to see, without curbing the lawful expression of others. Safety measures should focus on supporting agency and reducing risk—not eliminating sexual expression altogether.

Ultimately, both harm prevention and freedom of expression are best served by governance frameworks grounded in evidence, context, and accountability. Overbroad or opaque restrictions do not make people safer—they erode trust and obscure what truly works. By committing to transparent, proportionate, and empirically validated approaches, platforms and regulators can meaningfully reduce harm while upholding the diversity and dignity of global online communities.

Principle 6: Evaluate Impact on Marginalized Individuals and Communities

Poorly conceptualized and overly broad policies impact marginalized individuals and communities the most. Responsible and fair policies can only be created when those who are most likely to face erasure or restrictions, including survivors, sex workers, members of the LGBTQ community, and racial or ethnic minorities, are meaningfully involved in policy development. Policy changes should be evaluated for their impact on these communities.

This principle recognizes that broad safety policies can have disproportionate effects on those already facing structural inequality. What is framed as “protection” may result in erasure, exclusion, or criminalization—especially when policies are developed without the input of affected communities.

This is particularly true for survivors of violence, sex workers, LGBTQ+ people, Indigenous communities, people with disabilities, and racial or ethnic minorities. These groups often rely on digital spaces to share knowledge, organize for justice, or access resources unavailable to them elsewhere. When platforms or governments implement blunt or biased safety rules, these same users are often the first to be silenced or surveilled.

First and foremost, content-based rules must not be discriminatory. Enforcement should be grounded in behavior, not identity. Marginalized groups, including sex workers, queer communities, and trauma survivors, are disproportionately affected by opaque and discriminatory content moderation systems. Moderation policies should include safeguards against reinforcing bias or stigma.

In different societies, and to some extent globally, there are very strong stereotypes about sexuality and obscenity based on bodily features. Users on a platform may flag as obscene content involving a Black person or obese person, but not flag content of a very similar nature that involves an Asian person or someone else seen to be desirable or “allowed” to be sexual within a cultural context. Platforms should be responsible for ensuring that stereotyping and prejudice from their users do not corrupt their own content moderation policies, and it is the responsibility of the platforms (and other entities, including law enforcement) to understand this type of discrimination, staff with that expertise, and build awareness of what parity before content standards looks like in applied contexts.

But even when not explicitly discriminatory, minority communities may be disproportionately impacted in practice by content rules or by how they are applied. For example, while platform rules against bestiality content are uncontroversial in principle, in practice the way that these rules are interpreted and applied can impact disproportionately on queer communities that engage in pup play, furry fandom, and other niche interests, expressions, and practices.

Impact assessments must therefore be a central component of any content governance or online safety initiative, especially when deploying automated tools or participating in

industry-wide blocklists and hash databases. These assessments should be intersectional, disaggregated, and participatory—going beyond compliance checklists to ask: Who is helped by this policy? Who is harmed? Who was at the table when it was made?

Meaningful involvement of marginalized communities is not just a best practice—it is a requirement for fairness. Without it, digital safety frameworks risk replicating the very injustices they claim to prevent.

Conclusion

The principles outlined in this paper are not only vital for survivors of sexual violence and for those whose creative expression is at risk of being unfairly censored; they benefit every user who engages with digital spaces. Clear, evidence-based policies protect the integrity of online communities by ensuring that expression is not unjustly restricted and that survivors' experiences are neither trivialized nor co-opted. Transparent processes and accessible avenues for communication with platform teams foster trust, reduce frustration, and empower individuals to navigate digital environments more safely and confidently.

When resources are allocated to the prevention and investigation of real sexual abuse—rather than wasted on the pursuit of lawful, consensual, or fictional content—everyone benefits. Survivors gain stronger support systems, would-be victims are better protected, and communities can focus on building cultures of consent and respect. Meanwhile, companies that take a thoughtful, rights-based approach to content moderation see greater user engagement, reduced reputational risk, and fewer public controversies. By contrast, the fundamentally flawed and inconsistent practices that dominate the industry today serve little purpose beyond offering quick soundbites for politicians and sensational headlines for media outlets.

These principles represent an opportunity to move beyond rhetoric and towards meaningful action. They can serve as a framework for policymakers, trust and safety professionals, researchers, and community advocates seeking to balance safety with freedom of expression. We invite you to reference them in your own work, share them with colleagues, and incorporate them into advocacy, policy development, and digital governance initiatives. By embedding these values into practice, we can build digital spaces that truly protect survivors, respect creative and cultural expression, and strengthen the foundations of a more just and inclusive online world.

Acknowledgments

The Drawing the Line principles are a collaborative project, initiated by the Center for Online Safety and Liberty (COSL), and developed by an Advisory Board consisting of the following members, who graciously contributed their time and expertise to the project in their personal capacities:

- Aurélie Petit, a PhD Candidate in the Film Studies department at Concordia University, Montréal, and Guest Editor for the Porn Studies journal Special Issue on Artificial Intelligence, Pornography, and Sex Work.
- Emma Shapiro, Independent expert, and Editor-At-Large of Don't Delete Art , a project advocating for artists facing censorship online.
- Ira Ellman, Distinguished Affiliated Scholar, Center for the Study of Law and Society, University of California, Berkeley and Charles J. Merriam Distinguished Professor of Law and Affiliate Professor of Psychology, Emeritus, Arizona State University.
- Masayuki Hatta, associate professor of Economics and Management at Surugadai University, Japan and visiting researcher at the Center for Global Communications (GLOCOM), International University of Japan.
- Michael McGrady Jr, Journalist in 1st Amendment, sex work, LGBTQ+ rights, ethics and compliance (until November 2025).
- Ashley Remminga, a graduate researcher in Global Cultures and Languages at the University of Tasmania exploring queer and, in particular, transgender and gender-diverse participation and engagement within popular culture and fandom.
- Zora Rush, a linguist, works on Responsible AI at Microsoft, where she develops assessment systems for harmful content generation in LLM products, including sexual content.

This background paper was written by Jeremy Malcolm, Chair of COSL, and Noor Afrose, Project Coordinator and Principal Researcher for the Drawing the Line Project, with contributions by COSL Advisors Emma Shapiro and Zora Rush. This paper expresses the opinions of its authors only, and does not necessarily represent a consensus of all of those who have contributed towards or endorsed the Drawing the Line Principles.

Bibliography

Af viceland. (2006, May 22). Ten Years Ago in Vice—Daddy's Little Slut. *VICE*.

<https://www.vice.com/da/article/ten-years-ago-in-vice-daddys-little-slut-7/>

Beazley, J. (2025, April 15). Sydney author arrested over 'pedophilia' book. *The Guardian*.

<https://www.news.com.au/national/nsw-act/crime/sydney-author-lauren-tesolinmastrosa-arrested-over-pedophilia-book/news-story/5babb82438d7adc5ca699c877b07641a>

Bischoff, P. (2021, February 19). The rising tide of child abuse content on social media.

Comparitech.

<https://www.comparitech.com/blog/vpn-privacy/child-abuse-online-statistics/>

Carter, R. F. (2020). Author and Publisher of "Hansel et Gretel" Acquitted in Quebec Court.

Freedom to Read.

<https://www.freedomtoread.ca/newsbytes/author-and-publisher-of-hansel-et-gretel-acquitted-in-quebec-court/>

ECPAT. (2025). *Terminology Guidelines for the Protection of Children from Sexual Exploitation and Abuse* (Second Edition).

<https://ecpat.org/wp-content/uploads/2025/04/Second-Edition-Terminology-Guidelines-final.pdf>

Farokhmanesh, M. (2025, July 24). *Steam and Itch.io Are Pulling 'Porn' Games. Critics Say It's a Slippery Slope to More Censorship* | WIRED. WIRED.

<https://www.wired.com/story/steam-itchio-are-pulling-porn-games-censorship/>

franceinfo & AFP. (2025, February 13). L'auteur de BD Bastien Vivès renvoyé devant la justice pour fixation et diffusion d'images à caractère pornographique de mineurs. *Franceinfo*.

https://www.francetvinfo.fr/culture/bd/l-auteur-de-bd-bastien-vives-renvoie-devant-la-justice-pour-fixation-et-diffusion-d-image-a-caractere-pornographique-de-mineurs_7072455.html

Gazeta, N. (2019, December 2). *Russian trans woman sentenced to likely fatal three years in prison for posting manga on social media*. Meduza.

<https://meduza.io/en/feature/2019/12/02/russian-trans-woman-sentenced-to-potentially-fatal-three-years-in-prison-for-posting-manga-on-social-media>

GLAAD. (2025). *2025 Social Media Safety Index*.

https://assets.glaad.org/m/346d7b38bb818f6d/original/2025-Social-Media-Safety-Index.pdf?_gl=1*1y8at95*_gcl_au*NTY5Njl5NjI0LjE3NTMzOTk4OTk.

Hixenbaugh, M., Kingkade, T., & Friedman, S. (2024, July 23). *Inside the two-year fight to bring charges against school librarians in Granbury, Texas*.

<https://www.nbcnews.com/news/us-news/school-librarians-banned-books-investigation-texas-rcna161444>

Lanzarote Committee. (2024). *Declaration on protecting children against sexual exploitation and*

sexual abuse facilitated by emerging technologies.

<https://rm.coe.int/declaration-on-protecting-children-against-sexual-exploitation-and-sex/1680b25a78>

Letourneau, E. (2022, March 23). We're still not spending enough to prevent child sexual abuse.

The Hill.

<https://thehill.com/opinion/criminal-justice/599189-were-still-not-spending-enough-to-prevent-child-sexual-abuse/>

Lievesley, R., Harper, C. A., Woodward, E., & Tenbergen, G. (2023). Fantasy Sexual Material Use by People with Attractions to Children. *Current Psychiatry Reports*, 25(9), 395–404.

<https://doi.org/10.1007/s11920-023-01435-7>

Mackinnon, C. A. (1989). Sexuality, Pornography, and Method: "Pleasure under Patriarchy.

Ethics, 99(2), 314–346.

Malcolm, J. (2019). *Internet Reporting Hotlines Are Censoring Art as Child Sexual Abuse*. Circle ID.

https://circleid.com/posts/20191211_internet_reporting_hotlines_censoring_art_as_child_sexual_abuse

Malcolm, J. (2024, November 28). Australia Versus Human Rights Online. *Jeremy Malcolm*.

<https://jere.my/australia-versus-human-rights-online/>

Noble, S. U. (2018). *Algorithms of Oppression*. NYU Press.

<https://nyupress.org/9781479837243/algorithms-of-oppression/>

Petit, A. (2025). The limits of “zero tolerance” policies for animated pornographic media. *Porn Studies*. <https://doi.org/10.1080/23268743.2025.2491506>

Rouse, A., Jenkinson, E., & Warner, C. (2023). The use of “art” as a resource in recovery from the impact of sexual abuse in childhood: A qualitative systematic review. *Arts & Health*, 15(1), 86–109. <https://doi.org/10.1080/17533015.2022.2034900>

Safe Online. (2025). *Safe Online Impact 2024*. <https://safeonline.global/ar-2024/>

- Sandí, R. (2019, May 18). Joven de 17 años difundió por Internet 146 dibujos pornográficos de niños. *La Teja*.
<https://www.lateja.cr/sucesos/video-joven-de-17-anos-difundio-por-internet-14/66NV7C4YABAOPKTAU3YJ43BC2M/story/>
- SHG. (2021, July 16). The Most Repugnant Man. *Simple Justice*.
<https://blog.simplejustice.us/2021/07/16/the-most-repugnant-man/>
- Strossen, N. (2000). *Defending Pornography: Free Speech, Sex, and the Fight for Women's Rights*. NYU Press.
- Swan, R. (2023, October 6). *Santa Rosa man sentenced to 33 years for operating obscene website*.
<https://www.sfchronicle.com/crime/article/santa-rosa-man-sentenced-33-years-operating-18410898.php>
- Thorn & All Tech is Human. (2024). *Safety by Design for Generative AI: Preventing Child Sexual Abuse*. <https://info.thorn.org/hubfs/thorn-safety-by-design-for-generative-AI.pdf>
- UN Committee on the Rights of the Child. (2019, September 10). *Guidelines regarding the implementation of the Optional Protocol to the Convention on the Rights of the Child on the sale of children, child prostitution and child pornography*.
https://www.ohchr.org/sites/default/files/Documents/HRBodies/CRC/CRC.C.156_OPSC_Guidelines.pdf
- Walinchus, L., Smyser, K., & Murphy, J. (2025, January 10). *Three major takeaways we learned from 'Dismissed,' NBC News' investigation into how few violent sex crimes end in a conviction*. <https://www.nbcnews.com/specials/sex-assault-convictions/>
- Walker, A. (2023). Transphobic discourse and moral panic convergence: A content analysis of my hate mail. *Criminology*, 61(4), 994–1021. <https://doi.org/10.1111/1745-9125.12355>
- Ward, P. R. (2008, August 7). Writer's "monsters" lead to obscenity sentence. *Pittsburgh Post-Gazette*.

<https://www.post-gazette.com/local/neighborhoods/2008/08/08/Writer-s-monsters-lead-to-obscenity-sentence/stories/200808080169>

Wilde, T. (2025, April 10). Developer of “non-consensual sex” game withdraws it from Steam after it’s banned in the UK, Canada, and Australia: “We don’t intend to fight the whole world.” *PC Gamer*.

<https://www.pcgamer.com/games/developer-of-non-consensual-sex-game-removes-it-from-steam-entirely-after-its-banned-in-the-uk-canada-and-australia-we-dont-intend-to-fight-the-whole-world/>