AI & SOCIETY
https://doi.org/10.1007/s00146-026-03080-z

OPEN FORUM

Sexualized deepfakes as a socio-technical continuation of gendered power

Dana Mahr[1]

Received: 22 January 2026 / Accepted: 21 April 2026

Abstract
Sexualized deepfakes are frequently framed as a problem of deception, misinformation, or privacy. However, a growing body of scholarship has shown that the central harm of deepfake pornography does not primarily lie in viewers being deceived, but in the non-consensual creation and circulation of sexualized representations. Building on this insight, this article examines sexualized deepfakes as socio-technical mechanisms of gendered power. Drawing on science and technology studies and feminist media theory, it argues that deepfake pornography extends existing practices of image-based sexual abuse by weaponizing visual representation against women. The paper situates sexualized deepfakes within longer histories of visual domination and analyzes how digital platforms, AI infrastructures, and cultural norms amplify these harms. It shows that current governance approaches, including consent-based legal frameworks, victim-centered resilience strategies, and technical detection tools, address only partial dimensions of a fundamentally systemic problem. In response, the article proposes a consent-centered socio-technical framework for governing deepfake abuse that integrates platform accountability, legal reform, technical safeguards, and victim support. Understanding sexualized deepfakes as instruments of power rather than merely falsified media shifts the focus of governance from protecting authenticity toward protecting autonomy, dignity, and justice in the age of generative AI.

Keywords Sexualized deepfakes · Image-based sexual abuse · Gendered power · Visual coercion · Socio-technical governance · Generative AI

1 Introduction

In early 2026, a controversy erupted around X when its new AI chatbot "Grok" was found to be generating non-consensual sexualized images of women on demand. Within days, hundreds of these AI-altered photos (women digitally "undressed" or placed in pornographic scenarios) flooded the platform, prompting international alarm (Vicens and Satter 2026). Government officials in France and India announced investigations into X over the "sexual and sexist" deepfake images, calling the content manifestly illegal (Vicens and Satter 2026).

Deepfake technology (the AI-driven fabrication of realistic images, video, or audio) has sparked widespread concern about deception, misinformation, and the erosion of trust in visual evidence (Vaccari and Chadwick 2020; Giansiracusa 2021; Weikmann and Lecheler 2023). In public discourse and much early policy debate, deepfakes have often been framed primarily as threats to truth, democracy, and privacy, exemplified by concerns about fake news videos and electoral disinformation (Pawelec 2022; Gosse and Burkell 2020; Yadlin-Segal and Oppenheim 2021). However, a striking fact received comparatively less sustained public attention until recently: the vast majority of deepfake content circulating online is pornographic and disproportionately targets women.

A 2023 survey by the tech blog Security Hero found that 98% of deepfake videos online were pornographic (Security Hero 2023), and recent scholarship likewise emphasizes that women are overwhelmingly the primary targets of such content (Lapointe et al. 2025; McGlynn and Toparlak 2025; Lazard et al. 2025). In other words, the deepfake phenomenon is overwhelmingly a story of gendered sexual exploitation rather than political deception. Media analyses have shown that news coverage has often prioritized deepfakes as

✉ Dana Mahr
dana.mahr@kit.edu

[1] Institute for Technology Assessment and Systems Analysis, Karlsruhe Institute of Technology, Karlsruhe, Germany

Published online: 05 May 2026

problems of misinformation, democracy, and public deception rather than as forms of gendered or sexualized abuse (Gosse and Burkell 2020; Yadlin-Segal and Oppenheim 2021). In this open-forum piece, I argue that sexualized deepfakes are not simply deceptive media but technologies of power.

By superimposing a person's face (typically a woman's) onto pornographic imagery without consent, often using the body of an actual performer (also without consent) as the base, deepfakes weaponize representation itself. They coerce and violate through visuals, exploiting both the power of the gaze and the networked amplification of digital platforms (Lazard et al. 2025). As feminist media theory reminds us, visual media have long been implicated in gendered power dynamics. From the invention of the camera onward, the act of capturing and displaying a woman's image has often been an exercise of dominance, the male gaze that renders women objects for consumption (Mulvey 1975). As Berger observed, "Men act and women appear […] Men look at women. Women watch themselves being looked at," (Berger 2008, p. 47) highlighting how women are positioned as objects of vision within patriarchal structures. Sexualized deepfakes exemplify an extreme, technologically augmented extension of this dynamic: they make women appear in sexual scenarios they never chose, for the gratification or humiliation of (mostly) male spectators. In doing so, these fakes enact what I term visual coercion (a form of image-based force and intimidation) as opposed to mere deception. The goal is often not primarily to convince viewers that the target actually performed sexual acts, but rather to assert power over the target by sexually exposing, humiliating, and degrading her likeness (McGlynn and Toparlak 2025). This constitutes a new mode of symbolic violence: a harm exercised through representations and meanings, reinforcing gendered domination without laying a physical hand on the victim.

Situating deepfakes as instruments of gendered power requires an interdisciplinary analytical lens. In this open-forum article, I draw on science and technology studies (STS) to examine how social norms, values, and disvalues, including misogyny, become embedded in technologies and infrastructures, on feminist media and communication theory to understand the patriarchal context of visual culture, and on AI ethics and law to critique current governance approaches. The aim is to articulate a socio-technical perspective on sexualized deepfakes that can inform both understanding and response. By socio-technical, I mean an approach that sees technology and society as deeply interwoven: deepfakes are products of algorithms and data *and* of social norms, platform ecosystems, legal regimes, and power relations. I treat sexualized deepfakes as both an ethical and political question: how should societies understand and govern emerging technologies that enact gendered harm? This paper therefore takes an argumentative and reflexive stance, mapping key conceptual tensions to guide future empirical and policy work.

In the next Sect. (2), I review how scholars have conceptualized the harms of sexualized deepfakes, tracing a shift toward socio-technical and gendered power frameworks. In the third Sect. (3), I situate deepfakes within theories of visuality, gender, and harm, exploring how acts of seeing and showing function as mechanisms of power, how deepfake pornography extends gender-based violence, and how it destabilizes ideas of privacy and autonomy. The fourth section places this phenomenon in historical context (4), connecting it to earlier forms of image-based sexual abuse and revealing continuities of patriarchal intent. In the fifth Sect. (5), I critique current responses (legal, technical, and normative) that focus on consent, resilience, or detection, arguing that these approaches often overlook the broader socio-technical systems that enable harm. Finally, in the sixth Sect. (6), I outline a more holistic framework that integrates ethical AI design, platform accountability, education, and cultural change to address the misogyny underpinning these abuses.

2 Literature review

The rapid proliferation of sexualized deepfakes has prompted increasing scholarly attention across disciplines, yet the conceptualization of their harms remains fragmented across disciplinary approaches (Godulla et al. 2021; Twomey et al. 2025). Early research predominantly framed deepfakes as problems of misinformation, authenticity, or individual privacy, reflecting the dominance of technical and legal paradigms (Chesney and Citron 2018a; Gregory 2022; Amerini et al. 2025). More recent work, however, calls for socio-technical approaches that situate deepfakes within broader systems of social power (Vasist and Krishnan 2023; Agyare 2025). This literature review traces a shift from narrow concerns with deception and detection toward an understanding of sexualized deepfakes as a form of gender-based violence, while identifying key limitations that motivate the focus of this paper.

Initial deepfake scholarship was largely driven by computer science and engineering, prioritizing detection technologies and regulatory safeguards. Within this framework, deepfakes were treated primarily as deceptive media requiring authentication or prevention, reflecting the priorities of computer science and engineering research (Kietzmann et al. 2020; Brooks 2021; Kerner and Risse 2021; Mustak et al. 2023). Closely related misinformation paradigms (shaped by anxieties about political manipulation and "truth decay") emphasized risks to democratic discourse and public trust. While influential,

these approaches tended to obscure the social distribution of harm. Although some studies acknowledged pornographic deepfakes, early research rarely examined who was targeted and why, effectively treating deepfakes as a generalized problem of media authenticity rather than a gendered practice of abuse (Kerner and Risse 2021).

While early commentary on deepfake pornography often emphasized the risks of deception and misinformation, philosophical scholarship has increasingly argued that the central wrong of sexualized deepfakes lies elsewhere. Recent philosophical scholarship has argued that the central wrong of sexualized deepfakes lies not in deception but in the non-consensual creation and circulation of sexualized representations (Öhman 2020; De Ruiter 2021; Story and Jenkins 2023; Lorca 2025). These analyses show that deepfake pornography constitutes a form of image-based sexual abuse, even when viewers recognize the content as fabricated. The present article builds on this line of critique. Rather than redefining the problem of sexualized deepfakes, it extends these arguments by explicitly situating them within broader feminist analyses of visual culture and gendered power. In particular, it examines how deepfake technologies operate as socio-technical mechanisms that reproduce and amplify long-standing patterns of patriarchal domination through digital infrastructures.

Legal scholarship has developed in parallel, conceptualizing deepfake harms through individual rights such as privacy, consent, defamation, and control over one's likeness (Judge and Korhani 2021; Kadri and West 2025). Sexualized deepfakes were framed as violations of sexual privacy and extensions of image-based abuse. While these analyses articulated important legal harms and remedies, they largely remained focused on individualized violations and case-by-case responses (Dunn 2024; Karagianni and Doh 2024). As a result, both technical and legal paradigms (despite their contributions) tended to conceptualize deepfakes as discrete incidents rather than as systemic phenomena embedded in unequal social relations.

More recent scholarship adopts socio-technical and feminist perspectives that explicitly link technological affordances to social structures (Karagianni and Doh 2024; Wagner and Blewer 2019; Burkell and Gosse 2019; Lazard et al. 2025). Rather than treating deepfakes as neutral tools gone awry, this work emphasizes how design choices, platform infrastructures, and cultural norms collectively enable non-consensual sexualization (Wagner and Blewer 2019; Burkell and Gosse 2019; Lazard et al. 2025). Deepfakes are increasingly understood as continuations of existing practices of visual domination, intensified by automation and scale (Kim 2022; Oscar 2023; Gockel 2024). Feminist scholars in particular frame sexualized deepfakes as a form of gender-based violence that undermines women's agency, dignity, and autonomy, while also producing collective harm by creating a climate of threat that constrains women's participation in digital spaces (Taylor 2023; Karagianni and Doh 2024).

However, insufficient analytical attention has been paid to power relations themselves, despite increasing recognition that women are disproportionately targeted. Although many studies now acknowledge that women are the primary targets of sexualized deepfakes, gender is often treated descriptively rather than analytically. Harm is frequently explained through surface-level factors such as visibility, popularity, or image availability, rather than through patriarchal power structures that normalize the objectification and sexual appropriation of women's bodies (Fredrickson and Roberts 1997, pp. 175–185). As Burkell and Gosse (2019) argue, focusing narrowly on individual harm risks obscuring the cultural tolerance for misogyny that makes deepfakes resonate as effective attacks. Moreover, existing frameworks rarely address how power operates intersectionally, leaving limited analysis of how intersecting forms of structural inequality (including racism, homophobia, and transphobia) shape patterns of targeting and harm. Without centering these dynamics, analyses of deepfake harm remain conceptually incomplete and politically muted (Burkell and Gosse 2019).

In sum, while the literature has moved toward recognizing sexualized deepfakes as socio-technical and gendered harms, it has yet to fully theorize them as practices of power. Addressing this gap requires a framework that explicitly foregrounds visuality, gender, and structural domination as central analytical categories.

3 Conceptual framing: visuality, gender, and harm

Deepfakes do not emerge in a social vacuum; they tap into a long history of visual practices and gender norms. To conceptually frame sexualized deepfakes, we must consider three interlocking dimensions: (1) visuality and power, (2) gender and the continuum of sexual harm, and (3) the nature of the harm inflicted (including violations of autonomy and identity). This section lays out each of these dimensions, drawing on theoretical insights that situate deepfake porn as a socio-cultural phenomenon rather than a purely technical anomaly.

3.1 Visuality and power

Historically, visual representation has often been associated with relations of power in many societies. Feminist media and cultural theory have long emphasized that acts of looking and being seen are embedded in social hierarchies rather than being neutral processes (Mulvey 1975; Berger 2008).

The act of looking is not neutral; it can objectify, control, or claim ownership over the one who is seen. Feminist scholars of media and film have long analyzed the "male gaze." Visual culture frequently positions men as active viewers and women as passive objects to be viewed (Mulvey 1975; Schroeder 1998). The gaze operates as a power relation, a way of looking that positions the viewer as dominant and the viewed as subordinate (Schroeder 1998). In her classic essay on cinema, Laura Mulvey argued that mainstream films are structured around a male viewer's gaze, turning women on screen into erotic objects for male pleasure (Mulvey 1975). John Berger, as mentioned, observed that women internalize this objectification, learning to watch themselves through men's eyes (Berger 2008). These insights highlight that visual representations are never just images; they are relations of power. This dynamic also resonates with Foucauldian analyses of visibility and disciplinary power, in which subjects internalize the gaze as a mechanism of control and surveillance (Foucault 1977, pp. 202–203). The camera, as soon as it was invented, became a tool that could "capture" subjects, often reflecting the shooter's social dominance (e.g., colonial photography, or early pornographic images overwhelmingly made by and for men).

Sexualized deepfakes exploit this power of visuality in novel but deeply gendered ways. By digitally transplanting a woman's face onto pornographic bodies, deepfakers assert a kind of visual control over her, effectively saying: *"I can make you appear however I want, for my purposes."* This is an act of what we can call visual coercion: using images to force a narrative or scenario onto someone's identity. Unlike traditional deception (where the primary victim might be the fooled viewer), here the primary victim is the person depicted, and the image is a means of coercing or punishing her (Wagner and Blewer 2019; Taylor 2023). The target loses agency over how she is seen, becoming a puppet in a visual performance orchestrated by others (Paris 2021). In many documented cases, perpetrators appear to use deepfake pornography as a means of humiliating, harassing, or threatening women. Research on image-based sexual abuse shows that such practices are frequently embedded in patterns of online harassment and gendered intimidation (Henry and Powell 2015; Viola and Voto 2023).

In documented cases, perpetrators use deepfake pornography as a means of humiliating, harassing, or threatening women. For instance, Indian journalist Rana Ayyub was targeted in 2018 with a fabricated pornographic video circulated as retaliation for her outspoken activism (India Today 2018). While few believed the video was real, it unleashed a torrent of sexualized abuse and rape threats meant to silence and humiliate her. Similarly, German author and journalist Patrizia Schlosser was recently targeted by sexual deepfake websites and users (Moore and Ko 2025). In both cases, the harm lay not in deception but in violation and public shaming: the weaponization of visibility itself. A deepfake can therefore be *true* in what it exposes about the perpetrators' intent to dominate and sexualize, even as it is *false* in content. This perspective shifts our understanding from deepfakes as "fake news" to deepfakes as extensions of the age-old power to sexually objectify and silence women through imagery.

3.2 Gender and the continuum of sexual violence

Feminist analyses caution against viewing harms like deepfake porn in isolation or as aberrations. Instead, they are part of a continuum of sexual violence that spans from seemingly "minor" violations (harassment, non-consensual imagery) to extreme ones (physical sexual assault) (Kelly 1988). All these practices feed into a culture that tolerates and reproduces gendered violence. Non-consensual pornography, whether "real" (e.g., leaked nudes, secretly recorded videos) or fabricated via deepfakes, occupies a space on this continuum. While no physical contact occurs, the emotional, reputational, and psychological injuries inflicted by image-based abuse can be profound. Survivors of "revenge porn" and deepfake porn report feelings of violation akin to sexual assault, including trauma symptoms like anxiety, depression, and PTSD (Piper and Dardis 2025). The concept of symbolic violence, coined by sociologist Pierre Bourdieu, is useful here (Bourdieu 2001). Symbolic violence refers to subtler forms of violence that operate through symbols, language, and representations to reinforce dominance. Sexualized deepfakes are precisely a form of symbolic violence: they use sexualized images as a weapon to demean and assert power over women, reinforcing a social message of female subjugation. The fact that these images are fake does not neutralize their impact. On the contrary, it is precisely their falseness that reveals the imbalance of power: perpetrators can fabricate a degrading fiction and forcibly graft it onto a woman's public identity, without her consent or control.

It is also crucial to note the gendered asymmetry in who is targeted and who perpetrates these acts. Multiple studies confirm that women and girls constitute the overwhelming majority of victims of non-consensual intimate imagery, whereas perpetrators are predominantly men. Deepfake porn is no exception (Viola and Voto 2023). In practice, the production and circulation of deepfake pornography have so far been heavily gendered. Creators are typically male users who take publicly available images of women (often celebrities but increasingly also private individuals) and use them to generate sexually explicit content aimed primarily at male audiences (Öhman 2020; Lapointe et al. 2025; Viola and Voto 2023). It is a gendered abuse of technology, reflecting broader patterns of digital patriarchy where new tools are used to extend old sexist practices.

One telling example is the now-defunct DeepNude app: it allowed users to input a photo of a woman and receive a completely fabricated nude version of her. The app's algorithm was trained exclusively on images of female bodies, meaning it could only realistically "strip" *women*, not men (Cole 2019).

This design was not an accidental quirk but reflected the assumptions embedded in both the development process and the anticipated user base. The application was built to generate nude images of women, not men, because developers assumed a predominantly male audience seeking sexualized images of female bodies. In this sense, the design reproduced existing patterns of gendered objectification rather than introducing a neutral technological capability. As scholars of science and technology studies emphasize, technologies do not emerge in a social vacuum; they incorporate the values, expectations, and biases present in the societies and communities that produce them (Jasanoff 2010). The DeepNude case therefore illustrates how generative AI systems can encode and amplify misogynistic norms already present in digital culture. Rather than simply reflecting a vague "societal context," the system operationalized specific assumptions about who would be viewed, who would do the viewing, and whose bodies were considered legitimate objects of visual access.

Understanding deepfake porn as an extension of patriarchal power also invites us to consider how it intersects with other axes of oppression. For instance, women in public life (such as women politicians, journalists, or activists) are frequent targets, suggesting deepfakes are used to police gender norms and punish women who are visible or outspoken in male-dominated arenas. The sexualized defamation of female political figures via deepfakes (as in cases involving former U.S. Vice President Kamala Harris or Italian Prime Minister Giorgia Meloni) has been widely interpreted as an attempt to discredit and silence women in public life by invoking familiar sexist tropes, such as portraying female leaders as sexually "impure" or morally unfit for authority (Krook 2020). In this sense, sexualized deepfakes function less as mere prurient content than as a form of targeted gendered harassment. By attaching explicit imagery to women who occupy positions of political authority, such attacks draw on long-standing cultural narratives that police women's sexuality and undermine their credibility in public roles. The effect is not simply individual humiliation but the reinforcement of structural barriers to women's participation in political and professional life. It is a digital extension of practices that have long been used to keep women "in their place" through sexual shaming and intimidation.

3.3 Nature of the harm: autonomy and self-determination

What exactly is violated when someone is the victim of a sexual deepfake? One way to articulate this harm is through the concept of informational self-determination, the idea that individuals should retain meaningful control over how personal information about them is collected, processed, and represented.

The concept, originally developed in German constitutional law and later influencing European data protection frameworks (Rouvroy and Poullet 2009, pp. 57–59), has become foundational to contemporary privacy regimes. It underpins key instruments such as Article 8 of the EU Charter of Fundamental Rights and the General Data Protection Regulation (GDPR), both of which recognize individuals' rights to control their personal data and the digital representation of their identity.

In recent years, several jurisdictions have begun extending these principles to address non-consensual intimate imagery, including AI-generated content. For example, the United Kingdom's Online Safety legislation and related reforms now criminalize certain forms of non-consensual deepfake pornography, while several U.S. federal and state initiatives (including proposals such as the "Take It Down Act") seek to regulate the distribution of AI-generated intimate images. These developments indicate an emerging legal recognition that the harm of deepfakes lies not only in deception but in violations of autonomy, identity, and dignity.

Deepfakes profoundly undermine this autonomy, identity, and dignity: they wrest control of one's likeness and graft it onto situations one did not choose. Recent philosophical work on digital duplicates further highlights the implications of this loss of control. Danaher and Nyholm argue that AI systems increasingly enable the large-scale creation of digital replicas of individuals, raising concerns about the erosion of scarcity and control over personal identity in digital environments (Danaher and Nyholm 2024, 2025). In the context of sexualized deepfakes, this problem becomes especially acute: the technology allows unlimited sexualized reproductions of a person's likeness without their knowledge or consent.

In effect, deepfakes hijack one's image (a core aspect of personal identity) and insert it into a false (and often sexually explicit) narrative. This is a violation of privacy, but it is more than that. Legal scholar Danielle Citron describes such attacks as violations of sexual privacy, the right to have control over how (and whether) images of one's body are shared and portrayed (especially in sexual contexts) (Citron 2019). The non-consensual creation or sharing of intimate imagery is, in Citron's view, a form of sexual abuse that robs the individual of dignity and security in their own persona (Citron 2019).

Across jurisdictions, lawmakers are increasingly recognizing that synthetic media can inflict the same harms as real non-consensual pornography and are gradually adapting existing legal frameworks to address AI-generated intimate imagery. The harm is not tied to the factual accuracy of the image, but to the imposition of a sexual representation on someone without consent.

Empirical research on image-based sexual abuse and survivor testimonies (Henry and Powell 2015; McGlynn et al. 2017; Piper and Dardis 2025) highlights the profound psychological and social harms associated with non-consensual intimate imagery. Studies show that victims frequently report feelings of violation, loss of agency, and lasting damage to their sense of identity and reputation, with many experiencing symptoms consistent with trauma, including anxiety, depression, and post-traumatic stress (Piper and Dardis 2025; McGlynn et al. 2017). For example, Piper and Dardis (2025) find that many victims of non-consensual image dissemination report long-term emotional distress, social withdrawal, and difficulties re-establishing trust in personal and professional relationships. In this sense, the harm is not limited to the existence of a fabricated image but extends to the social meaning attached to it and the ongoing loss of control over one's public identity (Piper and Dardis 2025).

Researchers have described this as a "silencing effect," in which victims withdraw from public or online participation out of fear, shame, or anticipated harassment (Henry and Powell 2015; McGlynn et al. 2017). Rather than engaging openly in digital spaces, survivors often limit their visibility or retreat from public discourse altogether. Victims also suffer practical harms: some lose jobs or find their career prospects damaged because employers or colleagues encounter them in a (fake) sexualized context (Henry and Powell 2015; Citron 2022). Others face ruptures in personal relationships (Henry and Powell 2015). Even when friends or family believe the content is fake, the mere knowledge that such images exist can cause immense distress and social alienation (Citron 2022). Another key aspect of the harm is reputation and credibility: especially for women in fields like politics, academia, or journalism, a sexualized deepfake is used to undermine their credibility ("she's not a serious person, she's a sexual object") (Citron 2022). This type of character assassination via sexualization has a long sexist history; deepfakes provide a new medium to execute it at scale.

It bears emphasizing that these harms are magnified by the networked, persistent nature of digital media. Once a deepfake is released online, it can be widely distributed (across multiple platforms, often anonymously) and difficult to eradicate. The victim faces the Sisyphean task of trying to get each copy or link taken down, often needing to report content on sites that may or may not respond promptly. The platform infrastructure and visibility of content therefore become directly relevant to the impact of deepfakes. For now, suffice it to say that the viral visibility of a deepfake can itself be a source of harm: the knowledge that potentially thousands or millions of strangers have seen the fabricated pornographic image/video of oneself is a psychological burden that is hard to quantify. It is a form of mass public shaming or exposure. And unlike ephemeral gossip, digital artifacts can resurface repeatedly, renewing the trauma. This permanence and searchability of content in the platform age means the harm of a deepfake can be "evergreen," each new view or share re-victimizes the person. The stakes for personal security and dignity in this context are thus extremely high.

4 Sexualized deepfakes as a continuation of visual violence

Sexualized deepfakes did not invent image-based sexual abuse; they extend a long continuum of visual sexual violence that predates the digital era. To understand their significance, we must see how images have long been weaponized against women and how deepfakes both reproduce and transform those practices. This historical and sociotechnical lens reveals that deepfakes are not a "tech problem" but a digital evolution of enduring gendered power relations.

4.1 From revenge porn to deepfakes

The 2010s marked the rise of "revenge porn," the non-consensual sharing of intimate images, often by ex-partners seeking revenge. Alongside it, crudely photoshopped images circulated to sexualize or humiliate women. These forms led scholars to adopt the broader term "image-based sexual abuse" (IBSA), capturing the use of images themselves as tools of domination (Henry and Powell 2015; McGlynn et al. 2017). Deepfakes escalate this logic: perpetrators no longer require authentic intimate material. As Kugler and Pace (2021) emphasize, "using only a series of images of a person's face and publicly available software, it is now possible to insert the person's likeness into a video and show them saying or doing almost anything, (p. 611)" dramatically lowering the threshold for victimization. The result is a chilling sense of vulnerability: any woman, public or private, can be digitally stripped and exposed without ever having participated in such acts. As legal scholars Bobby Chesney and Danielle Citron (2019) argue, "Deep-fake sex videos can transform rape threats into a terrifying virtual reality. They send the message that victims can be sexually abused at whim" (p. 1773).

4.2 Visual violence, old and new

While the technology is novel, the underlying logic is familiar. Across history, women's images have been weaponized for humiliation, from medieval shame paintings to schoolyard "slut-shaming" collages (Ringrose and Harvey 2015; Milani 2016). In every case, the image serves as a public spectacle that disciplines and degrades. Deepfakes amplify this by making the fabricated spectacle seamless and infinitely shareable. The realism is secondary; what matters is the power to fantasize and circulate a woman's image as if she were available for sexual consumption. Whether or not viewers believe the video is real, the act reasserts the same patriarchal dynamic: control over women's visibility and sexuality.

4.3 Platforms and infrastructures of visibility

Deepfake pornography thrives on platform infrastructures designed for virality, scalability, and anonymity. The phenomenon gained traction in 2017 when a Reddit user began posting AI-generated celebrity porn, spawning a "r/deepfakes" community that quickly amassed over 100,000 followers before being banned, only for the content to migrate elsewhere (Karen 2019). By 2018, specialized deepfake porn websites had emerged, attracting over 134 million views within 18 months (Chesney and Citron 2018b). The architecture of digital platforms actively enables this scale of harm: search algorithms amplify visibility, hosting services provide infrastructure, and monetization systems sustain an economy around abuse.

Mainstream platforms such as Pornhub, Reddit, and Twitter (now X) eventually banned non-consensual deepfakes under public pressure, but enforcement remains inconsistent and difficult to assess. Since Elon Musk's acquisition of Twitter, researchers and journalists have noted significant changes to the platform's transparency and moderation infrastructure. For example, the restriction of API access and other policy changes have made it considerably harder for external researchers to monitor harmful content and moderation practices on the platform. Researchers have also noted that changes to X's API access policies have significantly reduced transparency for external researchers and watchdog organizations. As Elizabeth Blakey (2024) observes, the platform's new restrictions make it harder to see what is happening within its information ecosystem. While the precise effects on moderation remain contested, such reductions in transparency raise concerns about the capacity of platforms to detect and respond to abusive synthetic media at scale.

At the same time, automated moderation systems continue to struggle with non-consensual intimate imagery. While detection tools can identify certain forms of manipulated or explicit content, they remain fundamentally limited in assessing contextual factors such as consent, intent, and harm. As LeeYouk and Seering (2026) emphasize, "harm arises from the absence of consent rather than the sexual nature of the content itself," (p. 2) making it difficult for detection systems to infer illegality from visual features alone. Moreover, current moderation approaches are largely reactive and platform-bound, which limits their ability to effectively address the persistence and cross-platform re-circulation of such material. As a result, victims frequently bear the burden of identifying and reporting abusive content themselves, navigating slow or inconsistent reporting systems across multiple sites. In fact, X (under Elon Musk's ownership) has loosened safeguards, with the integration of its AI chatbot *Grok* raising new concerns about how generative tools might further normalize or facilitate such content. Automated filters miss much of the content, and the burden of reporting still falls on victims who often remain unaware of their exploitation (Pavón Pérez et al. 2026). Meanwhile, dedicated deepfake marketplaces continue to profit, offering "custom orders" and bounties for fake porn of specific women. This digital infrastructure ensures that visibility itself becomes complicit in the abuse.

4.4 Symbolic continuities

Deepfakes replicate older patterns of patriarchal control, particularly entitlement to women's bodies and the use of sexual shaming as social discipline. In appropriating a woman's face for pornographic display, perpetrators enact the same denial of agency that has long underpinned rape culture. Feminist scholars of image-based sexual abuse and digital sexual violence (e.g., McGlynn et al. 2017; Taylor 2023) describe this as a technological extension of sexual entitlement: the perceived right to use a woman's image without her consent. Even when viewers know a deepfake is fabricated, the stigma lingers (Citron 2022). A single video can permanently taint a woman's reputation in societies that still police female sexuality harshly. Meanwhile, perpetrators remain largely anonymous and unpunished, rewarded by online communities that normalize degradation as entertainment. In sum, sexualized deepfakes are not an aberration of AI but a continuation of visual violence. They merge historical misogyny with digital affordances, scaling old harms through new infrastructures. Recognizing these continuities underscores the need for responses that go beyond detection or takedown, responses that confront the cultural, economic, and platform conditions sustaining this abuse. The next section turns to governance: how institutions have responded, and why those efforts remain insufficient.

5 Governance and the limits of consent and resilience

As sexualized deepfakes have moved from a niche phenomenon to a recognized social harm, a range of governance responses has emerged across legal, technical, and social domains (McGlynn and Toparlak 2025; Henry and Powell 2015; Kietzmann et al. 2020). These can be broadly grouped into three approaches: legal measures centered on consent and liability (McGlynn and Toparlak 2025), individual and social measures emphasizing resilience and support (Henry and Powell 2015), and technical measures focused on detection and content moderation (Kietzmann et al. 2020). While each approach contributes something, none is sufficient on its own. Taken in isolation, all three tend to frame the problem too narrowly (either as a matter of individual consent, individual coping, or technical deception) thereby overlooking the broader socio-technical systems and power relations that enable and sustain the harm.

5.1 Legal approaches: consent, liability, and their limits

Legal responses typically frame deepfake pornography as a violation of consent and seek to criminalize or civilly sanction its creation and distribution. Several jurisdictions have moved in this direction, including U.S. states that have expanded "revenge porn" laws to cover synthetic media and the UK's 2023 legislation explicitly banning deepfake pornography (Dunn 2024). These measures are normatively important: they signal social condemnation, recognize victims' rights, and offer avenues for redress.

Recent legal developments further underscore the growing recognition of deepfake abuse as a serious form of harm. Across jurisdictions, lawmakers have begun to adapt existing frameworks and introduce new offenses targeting the creation and distribution of non-consensual synthetic sexual imagery, reflecting an emerging consensus that such practices warrant legal sanction (Diamantis et al. 2026). At the same time, early enforcement efforts reveal the practical limits of these approaches. Empirical and legal analyses suggest that successful prosecutions remain relatively rare, in part due to persistent challenges in identifying perpetrators, establishing jurisdiction across platforms, and responding to the scale and speed at which such content can be produced and disseminated.

Yet consent-focused legal frameworks face significant limitations. Enforcement is difficult across jurisdictions, enabling perpetrators and platforms to evade accountability through geographic arbitrage. Legal action is also burdensome for victims, often requiring identification of anonymous perpetrators and costly, retraumatizing proceedings (McGlynn et al. 2017; Sparks et al. 2023). Moreover, by framing harm primarily as an interpersonal rights violation, such laws risk obscuring the role of platforms in amplifying and monetizing abusive content. Even emerging efforts to impose platform liability raise difficult questions about definition, enforcement, and unintended effects on online speech. As a result, consent-based laws tend to be reactive, fragmented, and ill-suited to addressing the distributed and systemic nature of deepfake abuse.

5.2 Individual resilience and victim support

A second response emphasizes individual resilience, digital literacy, and victim support. Educational initiatives encourage skepticism toward visual media, while NGOs provide resources to help victims report content, seek legal remedies, and cope with psychological harm. These efforts are valuable, particularly in countering shame and affirming that responsibility lies with perpetrators, not victims (Paris and Donovan 2019).

However, resilience-oriented approaches carry structural limitations. They can unintentionally shift responsibility onto victims by implying that individuals must adapt to a hostile digital environment. In the context of deepfakes, advice centered on self-protection or reduced visibility is especially inadequate, since even innocuous images can be misused. Expecting people—particularly women—to withdraw from online or public life reinforces the very inequalities that deepfake abuse exploits. Support services, while essential, also reach only a small fraction of victims and function more as damage control than prevention. More broadly, public skepticism about the authenticity of images does little to mitigate harm when deepfakes are used not to deceive, but to harass, intimidate, or humiliate through exposure itself.

5.3 Technical detection and moderation

A third approach focuses on technological solutions, including automated deepfake detection and content moderation tools. While significant progress has been made, detection technologies face an inherent arms race: as detection improves, generation techniques adapt. Current tools remain imperfect, often producing probabilistic assessments that require human judgment, which is costly, slow, and context-dependent (Wang et al. 2024). Crucially, detection cannot resolve the core issue of consent. A system may identify whether content is synthetic, but not whether it is non-consensual, or whether real sexual content is being shared abusively.

Even when detection works, it rarely prevents rapid dissemination or re-uploading across platforms. More fundamentally, detection addresses falsity rather than violation. In sexualized deepfakes, the primary harm lies not in viewers being deceived, but in the coerced circulation of a sexualized representation itself. Knowing that an image is fake does not undo the humiliation, intimidation, or loss of dignity inflicted on the person depicted.

5.4 Synthesis

Taken together, these approaches reveal the limits of single-pronged governance strategies. Legal remedies affirm rights but struggle with enforcement and systemic accountability; resilience strategies support victims but risk individualizing responsibility; and technical solutions focus on deception while missing the core dynamics of coercion and power. Sexualized deepfakes thus constitute a wicked socio-technical problem, one that spans law, technology, culture, and gendered power relations. Addressing it requires moving beyond narrow framings toward integrated, structural responses that confront not only the symptoms of harm, but the conditions that make such harm possible.

6 Toward a socio-technical framework for deepfake governance

If sexualized deepfakes are best understood as a socio-technical continuation of gendered power, governance responses must move beyond narrow concerns with deception or technical detection. What is required instead are measures that directly address consent, accountability, and structural power relations in the digital ecosystem. This section outlines several policy directions that shift responsibility away from victims and toward the actors who design, deploy, and profit from deepfake technologies.

6.1 Platform accountability

First, platform infrastructures must assume greater responsibility for preventing and removing non-consensual deepfakes. Recent regulatory frameworks already provide a foundation for this approach. Under the European Union's Digital Services Act (2022), large platforms are required to identify and mitigate systemic risks associated with harmful content, including violations of privacy and dignity. Sexualized deepfakes clearly fall within this category. Explicitly classifying non-consensual deepfakes as illegal content would obligate platforms to detect, remove, and prevent the dissemination of such material more proactively.

Similar legal developments are emerging elsewhere. The United Kingdom's evolving regulatory framework on online safety and intimate image abuse criminalizes the creation and distribution of non-consensual intimate imagery, including AI-generated material (Dunn 2024). In the United States, federal initiatives such as the proposed *Take It Down Act* aim to criminalize the sharing of non-consensual intimate images and require platforms to respond rapidly to takedown requests. These developments indicate a growing recognition that synthetic media can produce harms equivalent to real image-based abuse.

However, regulation must also address platform incentives. Platforms that systematically fail to remove non-consensual deepfakes should face meaningful penalties, particularly where algorithms or monetization systems amplify the visibility of such content. Without stronger accountability mechanisms, the economic infrastructures surrounding deepfake pornography will continue to reward abuse.

6.2 Technical safeguards

Second, technical interventions should focus on consent verification and provenance tracking rather than detection alone. Research into watermarking and provenance verification for AI-generated media has advanced rapidly in recent years. For example, emerging watermarking and provenance verification systems can embed identifiers into synthetic images so that their artificial origin can be verified and traced (Wang et al. 2024). Such technologies could assist platforms in identifying manipulated media and preventing large-scale dissemination. At the same time, purely technical solutions face clear limitations. Detection systems remain locked in an arms race with generative models, while automated moderation struggles to determine whether explicit imagery is consensual or abusive. For this reason, proposals to embed unique biometric identifiers, sometimes described as "Image Rights ID" systems, raise significant practical and ethical concerns. As De Ruiter (2021) notes, visual likeness is not unique: unrelated individuals can resemble one another closely enough to produce false matches. Strict biometric identification systems therefore risk generating both false positives and new privacy risks.

Instead, a more flexible approach may involve opt-in consent registries in which individuals can voluntarily register images or likenesses as protected against non-consensual synthetic use. Such mechanisms align with emerging proposals for a "transparency-driven" and opt-in data right that would give individuals meaningful control over how identity-linked data is used in AI systems (Lim 2026). These registries could be paired with legal obligations requiring platforms and AI developers to respect such preferences. In parallel, researchers and regulators should invest in tools capable of auditing training datasets for the presence of personal data, thereby addressing ongoing concerns around

opacity, provenance, and accountability in generative AI systems.

6.3 Legal reforms

Third, legal frameworks must evolve to provide clearer remedies for victims of deepfake abuse. Many jurisdictions now criminalize the non-consensual distribution of intimate images, yet these laws often struggle to accommodate synthetic media (Chesney and Citron 2019; Dunn 2024). Victims may face uncertainty about whether existing statutes apply when the image itself is fabricated. One promising direction is to extend privacy and data protection rights to cover AI-generated representations of individuals. A right to request the deletion of AI-generated intimate imagery could operate similarly to existing "right to erasure" provisions under the GDPR. Such measures would recognize that individuals have legitimate claims over the digital use of their likeness even when the underlying image is artificially generated. In addition, civil remedies should be strengthened so that victims can seek damages from perpetrators, platforms, or intermediaries that knowingly facilitate deepfake abuse. Strengthening privacy torts, such as those recognized under statutes like California's Invasion of Privacy Act, could allow victims to pursue legal action even when perpetrators attempt to hide behind online anonymity.

6.4 Education and support

Finally, governance must address the social and cultural conditions that enable deepfake abuse. Public education initiatives should move beyond teaching audiences how to detect manipulated media and instead emphasize the ethics of digital consent and image sharing. Educational programs in schools and online platforms can integrate media literacy modules that address generative AI, consent, and image-based sexual abuse. Equally important is the development of support infrastructures for victims. Specialized counseling services, legal aid programs, and rapid-response takedown mechanisms should be funded and expanded in ways comparable to existing support systems for survivors of revenge pornography. Law-enforcement training is also essential: deepfake pornography should be treated not merely as a technical violation of platform policy but as a form of technology-facilitated sexual abuse.

6.5 Rebutting "Image Rights ID" proposals

Some commentators have suggested embedding biometric identifiers into images to prevent misuse of a person's likeness. While appealing in theory, such proposals encounter significant conceptual and technical obstacles. As De Ruiter (2021) observes, visual likeness is not uniquely identifiable in the same way as fingerprints or DNA. Even unrelated individuals can resemble one another sufficiently to trigger automated matches. Strict identity-matching systems could therefore generate false accusations while introducing new privacy risks through centralized biometric databases. For these reasons, governance strategies should prioritize right-based approaches rather than strict biometric enforcement. Mechanisms such as consent registries, dataset auditing, and strong legal penalties for non-consensual deepfake creation provide more flexible and proportionate safeguards.

6.6 Toward structural accountability

Taken together, these measures shift the burden of responsibility away from victims and toward the institutions that enable deepfake abuse. They align with feminist critiques of resilience-based approaches that expect individuals to adapt to hostile digital environments. Instead, consent-centered governance recognizes that protecting dignity in the age of generative AI requires structural accountability from platforms, developers, and regulators alike.

7 Conclusion

This article has argued that sexualized deepfakes should not primarily be understood as a problem of deception, but as a socio-technical extension of gendered power. By enabling the non-consensual fabrication and circulation of sexualized representations, deepfake technologies transform long-standing practices of image-based abuse into scalable digital infrastructures. The harm they produce lies less in misleading audiences than in imposing sexualized visibility on individuals without their consent.

Viewing deepfakes through this lens reveals that the problem cannot be solved by technical detection or individual coping strategies alone. Instead, sexualized deepfakes expose the broader socio-technical systems (platform architectures, economic incentives, cultural norms, and AI infrastructures) that enable and amplify gendered abuse online. Effective responses must therefore address these structural conditions rather than focusing narrowly on authenticity or misinformation. The consent-centered governance framework proposed here outlines one possible direction. By combining platform accountability, legal reform, technical safeguards, and victim support, such an approach shifts responsibility away from victims and toward the institutions that enable deepfake abuse.

Recent developments further underscore the urgency of this challenge. Reports of the widespread use of so-called "nudify" applications among teenagers, including in school contexts, indicate that the creation of non-consensual sexualized deepfakes is becoming increasingly normalized and

accessible (Burgess 2026). In these cases, often involving young girls as targets, the harm lies not in deception but in the imposition of sexualized visibility and social humiliation, mirroring the dynamics analyzed in this article.

Ultimately, the challenge posed by sexualized deepfakes is not only technical but political: it concerns how societies choose to govern emerging AI technologies in ways that protect autonomy, dignity, and equality. Addressing this challenge will require integrating feminist insights about power and visual culture into the design and governance of digital infrastructures themselves.

8 Conflict of interest

The authors declare no competing interests.

Author contribution The author solely conceived the study, developed the theoretical framework, conducted the analysis, and wrote the manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. Bundesministerium für Forschung, Technologie und Raumfahrt (Project: Deep-Prisma)

Data availability No datasets were generated or analysed during the current study.

8. References

Agyare P (2025) Gendered digital harms and youth ideological trajectories: Socioconomic challenges of digital platforms. SocioEcon Chall 9(3):77–96. https://doi.org/10.61093/sec.9(3).77-96.2025

Amerini I, Barni M, Battiato S, Bestagini P, Boato G, Bruni V, Caldelli R, De Natale F, De Nicola R, Guarnera L, Mandelli S, Majid T, Marcialis GL, Micheletto M, Montibeller A, Orrù G, Ortis A, Perazzo P, Puglisi G, Purnekar N, Salvi D, Tubaro S, Villari M, Vitulano D (2025) Deepfake media forensics: status and future challenges. J Imaging 11(3):73. https://doi.org/10.3390/jimaging11030073

Berger J (2008) Ways of seeing. Penguin, London

Blakey E (2024) The day data transparency died: how Twitter/X cut off access for social research. Contexts 23(2):30–35. https://doi.org/10.1177/15365042241252125

Bourdieu P (2001) Masculine domination. Stanford University Press, Stanford, California

Brooks CF (2021) Popular discourse around deepfakes and the interdisciplinary challenge of fake video distribution. Cyberpsychol Behav Soc Netw 24(3):159–163. https://doi.org/10.1089/cyber.2020.0183

Burgess M (2026) The global rise of AI "nudify" apps in schools. WIRED. Available at: https://www.wired.com/story/deepfake-nudify-schools-global-crisis. Accessed 16 April 2026

Burkell J, Gosse C (2019) Nothing new here Emphasizing the social and cultural context of deepfakes. FM. https://doi.org/10.5210/fm.v24i12.10287

Chesney R, Citron DK (2018) Deep fakes: a looming challenge for privacy, democracy, and national security. SSRN Journal. https://doi.org/10.2139/ssrn.3213954

Citron DK (2022) The fight for privacy: protecting dignity, identity and love in our digital age. Chatto and Windus, London

Citron DK (2019) Sexual Privacy. In: Yale Law Journal. https://yalelawjournal.org/article/sexual-privacy. Accessed 22 Jan 2026

Cole S (2019) This Horrifying App Undresses a Photo of Any Woman With a Single Click. In: VICE. https://www.vice.com/en/article/deepnude-app-creates-fake-nudes-of-any-woman/. Accessed 22 Jan 2026

Danaher J, Nyholm S (2024) Digital duplicates and the scarcity problem: might AI make us less scarce and therefore less valuable? Philos Technol 37(3):106. https://doi.org/10.1007/s13347-024-00795-z

Danaher J, Nyholm S (2025) The ethics of personalised digital duplicates: a minimally viable permissibility principle. AI Ethics 5(2):1703–1718. https://doi.org/10.1007/s43681-024-00513-7

De Ruiter A (2021) The distinct wrong of deepfakes. Philos Technol 34(4):1311–1332. https://doi.org/10.1007/s13347-021-00459-2

Diamantis ME, Sullivan S, Alshanetsky E (2026) Deepest fakes. George Wash L Rev 94:1–45. https://www.gwlr.org/wp-content/uploads/2026/02/94-Geo.-Wash.-L.-Rev.-1.pdf. Accessed 16 Apr 2026

Dunn S (2024) Legal Definitions of Intimate Images in the Age of Sexual Deepfakes and Generative AI. McGill LJ 395 69(4): https://lawjournal.mcgill.ca/article/legal-definitions-of-intimate-images-in-the-age-of-sexual-deepfakes-and-generative-ai/. Accessed 16 April 2026

Foucault M (1977) Discipline and punish: the birth of the prison. Pantheon Books, New York

Fredrickson BL, Roberts T-A (1997) Objectification theory. Toward understanding women's lived experiences and mental health risks. Psychol Women Q 21:173–206

Giansiracusa N (2021) Deepfake deception: what to trust when seeing is no longer believing. How algorithms create and prevent fake news. Apress, Berkeley, CA, pp 41–66

Gockel J (2024) Current state of pornographic deepfakes: A Science and Technology Studies perspective. MJLA. https://doi.org/10.26481/mjla.2024.v15.1005

Godulla A, Hoffmann CP, Seibert D (2021) Dealing with deepfakes – an interdisciplinary examination of the state of research and implications for communication studies. SCM 10(1):72–96. https://doi.org/10.5771/2192-4007-2021-1-72

Gosse C, Burkell J (2020) Politics and porn: how news media characterizes problems presented by deepfakes. Crit Stud Media Commun 37(5):497–511. https://doi.org/10.1080/15295036.2020.1832697

Gregory S (2022) Deepfakes, misinformation and disinformation and authenticity infrastructure responses: Impacts on frontline witnessing, distant witnessing, and civic journalism. Journalism 23(3):708–729. https://doi.org/10.1177/14648849211060644

Henry N, Powell A (2015) Beyond the 'sext': technology-facilitated sexual violence and harassment against adult women. Aust N Z

J Criminol 48(1):104–118. https://doi.org/10.1177/0004865814524218
Security Hero (2023) 2023 State Of Deepfakes: Realities, Threats, And Impact. https://www.securityhero.io/state-of-deepfakes/. Accessed 21 Jan 2026
Jasanoff S (ed) (2010) States of knowledge: the co-production of science and social order, transferred to digital print. Routledge, London
Judge EF, Korhani AM (2021) Deepfakes, counterfeits, and personality. SSRN J. https://doi.org/10.2139/ssrn.3893890
Kadri TE, West SR (2025) Deepfake torts: Emerging tort frameworks in U.S. deepfake regulation. J Tort Law 18(2):515–552. https://doi.org/10.1515/jtl-2025-0032
Karagianni A, Doh M (2024) A feminist legal analysis of non-consensual sexualized deepfakes: contextualizing its impact as AI-generated image-based violence under EU law. Porn Stud. https://doi.org/10.1080/23268743.2024.2408277
Karen H (2019) The biggest threat of deepfakes isn't the deepfakes themselves. In: MIT Technology Review. https://www.technologyreview.com/2019/10/10/132667/the-biggest-threat-of-deepfakes-isnt-the-deepfakes-themselves/. Accessed 22 Jan 2026
Kelly L (1988) Surviving sexual violence. Polity Press; B. Blackwell, Cambridge, UK: Oxford, UK
Kerner C, Risse M (2021) Beyond porn and discreditation: Epistemic promises and perils of deepfake technology in digital lifeworlds. Moral Philos Polit 8(1):81–108. https://doi.org/10.1515/mopp-2020-0024
Kietzmann J, Lee LW, McCarthy IP, Kietzmann TC (2020) Deepfakes: trick or treat? Bus Horiz 63(2):135–146. https://doi.org/10.1016/j.bushor.2019.11.006
Kim E (2022) On the depth of fakeness. Deep fakes, 1st edn. Routledge, London, pp 50–70
Krook ML (2020) Violence against Women in Politics, 1st edn. Oxford University Press
Kugler MB, Pace C (2021) Deepfake privacy: attitudes and regulation. Nw U L Rev 116(3):611–680. https://doi.org/10.2139/ssrn.3781968
Laffier J, Rehman A (2023) Deepfakes and harm to women. J Digit Law Policy 3(1):1–21. https://doi.org/10.51357/jdll.v3i1.218
Lapointe VA, Dubé S, Rukhlyadyev S, Kessai T, Lafortune D (2025) The present and future of adult entertainment: a content analysis of AI-generated pornography websites. Arch Sex Behav. https://doi.org/10.1007/s10508-025-03099-1
Lazard L, Capdevila R, Turley EL, Gilfoyle K, Stavropoulou N (2025) Deepfake Technology and Gender-Based Violence: A Scoping Review. Trauma Violence Abuse. https://doi.org/10.1177/15248380251384271
LeeYouk S, Seering J (2026) Deepfake, Real Harm: A Participatory Approach for Imagining Infrastructures to Combat Deepfake Sexual Abuse. In: Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '26). ACM, New York. https://doi.org/10.1145/3772318.3790902
Lim D (2026) The surprising virtues of heterogeneity: legal pluralism and the governance of generative AI. Singap J Leg Stud. https://doi.org/10.2139/ssrn.6336502
Lorca JG (2025) It's not porn, it's sexual abuse: a scoping review of sexual deepfakes public opinions, perpetration, and harms. Violence Gender 12(4):95–109. https://doi.org/10.1177/23267836251389595
Martin N (2021) Image-based sexual abuse and deepfakes: a survivor turned activist's perspective. In: Powell A, Flynn A, Sugiura L (eds) The Palgrave handbook of gendered violence and technology. Springer International Publishing, Cham, pp 55–72
McGlynn C, Toparlak RT (2025) The 'new voyeurism': criminalizing the creation of 'deepfake porn.' J Law Soc 52(2):204–228. https://doi.org/10.1111/jols.12527
McGlynn C, Rackley E, Houghton R (2017) Beyond 'revenge porn': the continuum of image-based sexual abuse. Fem Leg Stud 25(1):25–46. https://doi.org/10.1007/s10691-017-9343-2
Milani G (2016) The Ban and the Bag: How Defamatory Paintings Worked in Medieval Italy. In: Behrmann C (ed) Images of Shame. De Gruyter, pp 119–140
Moore A, Ko L (2025) 'It was about degrading someone completely': the story of Mr DeepFakes – the world's most notorious AI porn site. The Guardian
Mulvey L (1975) Visual pleasure and narrative cinema. Screen 16(3):6–18. https://doi.org/10.1093/screen/16.3.6
Mustak M, Salminen J, Mäntymäki M, Rahman A, Dwivedi YK (2023) Deepfakes: deceptions, mitigations, and opportunities. J Bus Res 154:113368. https://doi.org/10.1016/j.jbusres.2022.113368
Öhman C (2020) Introducing the pervert's dilemma: a contribution to the critique of deepfake pornography. Ethics Inf Technol 22(2):133–140. https://doi.org/10.1007/s10676-019-09522-1
Oscar S (2023) Curious spectatorship in the age of deepfakes. Digital Creativity 34(3):230–247. https://doi.org/10.1080/14626268.2023.2248964
Paris B (2021) Configuring fakes: digitized bodies, the politics of evidence, and agency. Soc Media Soc 7(4):20563051211062919. https://doi.org/10.1177/20563051211062919
Paris B, Donovan J (2019) Deepfakes and Cheap Fakes. In: Data & Society. https://datasociety.net/library/deepfakes-and-cheapfakes/. Accessed 22 Jan 2026
Pavón Pérez Á, Farrell T, De Kock C, Jurasz O (2026) Still unsafe: what's holding us back on online safety for women. AI Ethics 6:246. https://doi.org/10.1007/s43681-026-01097-0
Pawelec M (2022) Deepfakes and democracy (theory): how synthetic audio-visual media for disinformation and hate speech threaten core democratic functions. DISO 1(2):19. https://doi.org/10.1007/s44206-022-00010-6
Piper CM, Dardis CM (2025) "I Still Haven't Worked Through It": A Mixed Methods Examination of Social and Psychological Outcomes of Nonconsensual Distribution of Sexually Explicit Materials. J Fam Viol. https://doi.org/10.1007/s10896-025-01008-7
Ringrose J, Harvey L (2015) Boobs, back-off, six packs and bits: mediated body parts, gendered reward, and sexual shame in teens' sexting images. Continuum 29(2):205–217. https://doi.org/10.1080/10304312.2015.1022952
Rouvroy A, Poullet Y 2009, The right to informational self-determination and the value of self-development: reassessing the importance of privacy for democracy. in Reinventing Data Protection: Proceedings of the International Conference (Brussels, 12–13 October 2007). Springer, Dordrecht: 45–76.
Schroeder JE (1998) Consuming representation: A visual approach to consumer research. In: Sherry JF (ed) Servicescapes: the concept of place in contemporary markets. NTC Business Books, Chicago, pp 109–122
Sparks B, Stephens S, Trendell S (2023) Image-based sexual abuse: victim-perpetrator overlap and risk-related correlates of coerced sexting, non-consensual dissemination of intimate images, and cyberflashing. Comput Hum Behav 148:107879. https://doi.org/10.1016/j.chb.2023.107879
Story D, Jenkins R (2023) Deepfake pornography and the ethics of non-veridical representations. Philos Technol 36(3):56. https://doi.org/10.1007/s13347-023-00657-0
Taylor D (2023) Technologies of Women's (Sexual) Humiliation. In: Feminist Philosophy and Emerging Technologies. Routledge
India Today (2018) I was vomiting: Journalist Rana Ayyub reveals horrifying account of deepfake porn plot. https://www.indiatoday.in/

trending-news/story/journalist-rana-ayyub-deepfake-porn-13934 23-2018-11-21. Accessed 22 Jan 2026

Twomey J, Ching D, Aylett PM, Quayle M, Linehan C, Murphy G (2025) What is so deep about deepfakes? A multi-disciplinary thematic analysis of academic narratives about deepfake technology. IEEE Trans Technol Soc 6(1):64–79. https://doi.org/10.1109/TTS.2024.3493465

Vaccari C, Chadwick A (2020) Deepfakes and disinformation: exploring the impact of synthetic political video on deception, uncertainty, and trust in news. Soc Media Soc 6(1):2056305120903408. https://doi.org/10.1177/2056305120903408

Vasist PN, Krishnan S (2023) Engaging with deepfakes: a meta-synthesis from the perspective of social shaping of technology theory. INTR 33(5):1670–1726. https://doi.org/10.1108/INTR-06-2022-0465

Vicens AJ, Satter R (2026) Elon Musk's Grok AI floods X with sexualized photos of women and minors. Reuters. https://www.reuters.com/legal/litigation/grok-says-safeguard-lapses-led-images-minors-minimal-clothing-x-2026-01-02/ Accessed 16 April 2026

Viola M, Voto C (2023) Designed to abuse? Deepfakes and the non-consensual diffusion of intimate images. Synthese 201(1):30. https://doi.org/10.1007/s11229-022-04012-2

Wagner TL, Blewer A (2019) "The word real is no longer real": deepfakes, gender, and the challenges of AI-altered video. Open Inf Sci 3(1):32–46. https://doi.org/10.1515/opis-2019-0003

Wang T, Liao X, Chow KP, Lin X, Wang Y (2024) Deepfake detection: a comprehensive survey from the reliability perspective. ACM Comput Surv 57(3):58:1-58:35. https://doi.org/10.1145/3699710

Weikmann T, Lecheler S (2023) Visual disinformation in a digital age: a literature synthesis and research agenda. New Media Soc 25(12):3696–3713. https://doi.org/10.1177/14614448221141648

Yadlin-Segal A, Oppenheim Y (2021) Whose dystopia is it anyway? Deepfakes and social media regulation. Convergence 27(1):36–51. https://doi.org/10.1177/1354856520923963