

# FEMINIST AI ETHICS, DEEPPAKE PORNOGRAPHY, AND THE EXPANSION OF THE PHILIPPINE SAFE SPACES ACT

**Enrique Benjamin R. Fernando III**  
*University of Cambridge, England*

**Sofia Kisse C. Luna**  
*University of Sydney, Australia*

**Junno Alfonsus M. Salvanera**  
*University of the Philippines Diliman, Philippines*

*On 3 August 2019, Republic Act (R.A.) 11313, also known as the Philippine Safe Spaces Act or Bawal Bastos Law, came into effect in the Philippines. It aimed to supplement R.A. 7877 or the Anti-Sexual Harassment Act of 1995 by expanding the legal scope of the definition of “sexual harassment” to include online forms thereof. The problem, however, is that the rise of artificial intelligence over the last five years has created new threats that even the Safe Spaces Act has failed to account for. In this article, we focus on the specific problems that AI-driven deepfake pornography poses and argue that the Safe Spaces Act needs to be amended further to classify deepfake pornography as a form of sexual harassment. In Part I, we introduce the framework of liberal feminism and two of its goals in relation to pornography that provide the moral justification for expanding the Philippine Safe Spaces Act: (1) the rejection of objectification and (2) the protection of sexual autonomy. In Part II, we formulate two legal principles that cover a broad set of actions and norms that, in turn, provide the jurisprudential justification for expanding the Philippine Safe Spaces Act. In Part III, we apply our insights from Parts I and II to formulate recommendations rooted in Feminist AI ethics for more concrete amendments to relevant laws and AI-policy guidelines that aim to protect women. In Part IV, we respond to the objection that the regulation of deepfake pornography constitutes an assault on free speech.*

*Keywords: AI Ethics, Deepfake Pornography, Liberal Feminism, Philippine Safe Spaces Act, Philosophy of Law*

## INTRODUCTION

On 3 August 2019, Republic Act (R.A.) 11313, also known as the Safe Spaces Act or *Bawal Bastos Law*, came into effect in the Philippines. It aimed to supplement

R.A. 7877 or the Anti-Sexual Harassment Act of 1995 by expanding the legal scope of the definition of “sexual harassment” from primarily physical forms thereof to include online forms such as:

...acts that use information and communications technology in terrorizing and intimidating victims through physical, psychological, and emotional threats, unwanted sexual misogynistic, transphobic, homophobic and sexist remarks and comments online whether publicly or through direct and private messages, invasion of victim’s privacy through cyberstalking and incessant messaging, uploading and sharing without the consent of the victim, any form of media that contains photo, voice, or video with sexual content, any unauthorized recording and sharing of any of the victim’s photos, videos, or any information online, impersonating identities of victims online or posting lies about victims to harm their reputation, or filing, false abuse reports to online platforms to silence victims (Section 12, R.A. 11313).

The expansion of the Anti-Sexual Harassment Act was celebrated at the time as a necessary response to the various forms of sexual harassment that had become prevalent on social media. These include both active and passive cases of verbal sexual harassment, the non-consensual sharing of personal photos and videos, online forms of exploitation and intimidation, and sexualized bullying (Gonzales 2022, 1181). More recently, however, the rapid integration of artificial intelligence (AI) into everyday life in the last three years has bred new forms of sexual harassment, which have subsequently necessitated yet another amendment to stay relevant with the times. On 14 January 2025, the Senate approved Senate Bill 2897, which aims to amend R.A. 11313 by expanding its scope to include AI and other emerging technologies. According to its authors, among the objectives of Senate Bill 2897 (“An Act Amending Republic Act No. 11313 Otherwise Known as the ‘Safe Spaces Act’ and for Other Purposes”) was to address, among other things, the rise of “deepfake” pornographic videos that depict real-life individuals—most of whom are women—engaging in sexual acts by altering previously existing photographs or videos (Senate of the Philippines 2025). Deepfake pornographic videos are often so realistic that they are indistinguishable from real pornographic videos and given that they are often circulated without the subject’s consent, they often lead to severe forms of emotional, psychological, and reputational harm against women. The bill has therefore been presented as a necessary amendment to its predecessors, given their silence on AI and the unique threats that did not previously exist on more traditional online platforms.

The problem, however, is that there is still a relative lack of literature on the nature of deepfake pornography and whether it qualifies as a form of sexual harassment broadly construed. Even worse, in various legal systems, deepfake pornography is sometimes exempted from being covered under its scope for various reasons, such as the fact that it is potentially protected under the First Amendment or free speech laws (Barber 2023, 2; Bourgault 2025, 51). Legal technicalities also enable its spread. In the Philippines in particular, the Philippine National Police Anti-Cybercrime Group admitted to the difficulties of prosecuting cases of deepfake

pornography due to a lack of an explicit legal definition, while Gabriela Women's Party legal counsel pointed out the difficulties of having to authenticate it as a form of electronic evidence and proving the chain of custody (Mangalus 2024).

Against this backdrop, the primary objective of this paper is to examine the nature of deepfake pornography and to provide a philosophical justification for its inclusion under sexual harassment laws in the Philippines. To this end, the rest of this paper is divided into four parts. In Part I, we establish the liberal feminist framework, which will provide the moral and feminist standards according to which deepfake pornography must be rejected. More specifically, we lay out the liberal feminist objectives of (i) rejecting objectification and (ii) protecting sexual autonomy as the basis for justifying the expansion of the Philippine Safe Spaces Act. In Part II we explore how these liberal feminist principles can be applied to provide jurisprudential justification. This aims to put forth a sense of recognition on the creation and proliferation of deepfake pornography as a grave offense distinguishable from other online forms of sexual harassment that must be explicated within the expansion of the Philippine Safe Spaces Act. In Part III, we propose guidelines for protecting the underlying principles built on in the previous parts. Such guidelines include changes to legislation and to the way AI is trained and used. In Part IV, we respond to one of the strongest principled objections to our argument: that regulating deepfake technology violates free speech.

## **LIBERAL FEMINISM AND THE NATURE OF DEEPPAKE PORNOGRAPHY**

### **Liberal Feminism in the Philippines**

Liberal feminism broadly construed is the view that women are rational and autonomous individuals who are entitled to the free exercise of their basic liberties as well as equal treatment in all spheres of social life, especially in relation to men (Tong 2018, 11). Historically, it has attempted to advance its objectives through legal and political reforms. This includes fighting for positive liberties such as the right to expression, education, or vote, as well as positive equalities such as equal pay in the workforce, as well as representation in public office. It also includes negative liberties such as freedom from assault, constraint, or coercion, as well as negative equalities such as protection from gender-based discrimination. More recently, liberal feminists have turned the spotlight on sexual violence inflicted against women, such as harassment, rape, and restrictions placed upon their reproductive rights (Russell et al. 2024, 243-244). Their methods have ranged from organizing women's marches and protests on the streets (Martin and Smith 2020, 4) to participating in online activism such as the #MeToo movement (Trott 2020, 1125), emphasizing the female experience of sexual objectification, especially in cases of skewed power relations or of a lack of legal mechanisms for upholding accountability. While grassroots movements and citizen activism have evolved over time, their end goal of achieving reform through legal and political processes has generally remained constant.

Liberal feminism is likewise an increasingly popular feminist ideology in the Philippines that has achieved much social, political, and legal success. For the purposes

of this paper, we highlight two of its central objectives, especially in relation to female sexuality: (i) the rejection of sexual objectification, and (ii) the protection of sexual autonomy. We use these two objectives to frame our discussion of deepfake pornography as well as to provide the moral justification for classifying it as a crime under the Philippine Safe Spaces Act. We shall also use them to provide guidelines for developing a feminist AI ethics that is nuanced to the Philippine context.

(i) **Rejection of Sexual Objectification.** As a traditionally conservative, Catholic, and androcentric society, the Philippines continues to espouse patriarchal structures and norms. The structures include laws and institutions like schools, banks, and police enforcement whose operations or lack thereof systematically disadvantage women and deprive them of equal treatment relative to men, whereas the norms include cultural and social practices which relegate women to a lower status within their communities. They form the collective backdrop against which women are treated as objects whose value primarily derives from the sexual gratification they give men. For instance, whereas men are often the breadwinners of households who control finances and decision-making, women are often subordinated to domestic chores and child-rearing with little to no authority over family matters (Aguilar 1989, 543; Reyes-Espiritu 2023, 80). In such male-dominated households, husbands often view their wives as mere sex objects to be used after a long day's work, and wives who are forced into sexual relations may be legally prevented from suing their husbands on marital rape charges.<sup>1</sup> Less extremely but no less worryingly, Filipino women are continuously sexually objectified, harassed, and cat-called on the street due to the rampant use of sexually-charged advertisements and commercials in Philippine media. In more severe cases of sexual violence, women are still subject to victim-blaming by being portrayed as promiscuous, lewd vixens who were "asking for it" due to their manner of dress (Benonguil 2022, 69; Valdez et al. 2022, 1).

Liberal feminists in the Philippines have made tremendous advances in terms of reducing the sexual objectification that women experience today. This can be observed in the increase of legislation aiming to protect women from new forms of sexual violence in recent years. A good example is House Bill 8655, or the Expanded Anti-Violence Against Women and Their Children (E-VAWC), which amends the original Anti-VAWC of 2004 by adding electronic or information and communications technology-related violence under its scope, which previously included physical, sexual, and psychological forms thereof. Similarly, the Cybercrime Prevention Act of 2012—while controversial in some respects—was celebrated for being one of the first laws in the Philippines that criminalizes cybercrimes like child pornography that frequently target women. Finally, Republic Act 11313, or the Safe Spaces Act (Bawal Bastos Law) of 2019, is also considered to be the product of liberal feminists' efforts to weaken traditional patriarchal and misogynist structures that enable the objectification of women. On the political side, the Philippine Commission on Women, which generally espouses a liberal feminist character, continues to lobby for political reform. Among its achievements are introducing the Women's Priority Legislative Agenda which proposes bills and amendments to existing laws that promote women's interests that align with the Philippine Magna Carta of Women (RA 9710), which is the most comprehensive law for protecting women's rights in the Philippines, as well as the Gender Equality and Women's Empowerment (GEWE)

Plan 2019-2025 which provides a framework for gender equality and women's empowerment against various forms of discrimination.

(ii) **Protection of Sexual Autonomy.** This second liberal feminist objective is related to but conceptually different from the first. The sense of 'autonomy' we are interested in may be clarified by the distinction between procedural and substantive sexual autonomy (Miller 2022, 248-251). Someone is procedurally sexually autonomous based on whether she has control over *how* she has chosen to engage in some sexual activity, i.e., if she has engaged in sex free from any sort of constraint or coercion. This conception of sexual autonomy is also referred to as consensual minimalism, or 'thin' sexual autonomy, because the mere existence of voluntary informed agreement is a necessary and sufficient condition for someone to be procedurally sexually autonomous. There may, however, be cases wherein a woman is not fully autonomous in the genuine, robust sense, such as when a woman provides sexual favors for men in exchange for monetary compensation or favors. In such cases, women may appear to consent to sex on the surface, but they do so for the wrong reasons; that is, they would not have consented to sex under different circumstances. In other words, on the substantive conception of autonomy, consent must result from "self-determined behavior" (Valentiner 2021, 706), arise from the right desires and preferences, and enable human flourishing. Hence, on this more demanding view, consent is not a matter of a simple 'yes' or 'no'.<sup>2</sup> On both procedural and substantive conceptions, Filipino women have traditionally lacked autonomy over their sexuality. They have been conditioned to subscribe to conservative views of intimacy, including the notion that a woman's role is to pleasure her partner; the failure to do so makes her less of an ideal woman.

Liberal feminists in the Philippines have also made headway on this issue. Payongayong, for instance, claimed after studying women's ethical issues that arose from their intimate relationships, family lives, and involvement in feminism that a "reflexive-liberative" ethics is gradually emerging in female consciousness:

It is reflexive because feminists have choices and understand the choices they make. They either reject or accept mainstream/male-centered ethics based on whatever works for them and other people. Reflexive further means continuous self-examination where ethical decisions are based on changing contexts and situations, and in consideration of the greater good. It is liberative because the choice is supposed to be not only for the benefit of feminists as individuals or a group, but for humanity as a whole. The feminist as a moral agent determines the liberative aspect. Being a feminist means there is recognition of the fact that women suffer from inequality, and there is a conscious effort to change this situation (Payongayong 2020, 91-92).

In the context of sexual autonomy, our interpretation is that the "liberative" aspect corresponds to the procedural conception of whether women can choose the sexual activities they engage in, while the "reflexive" aspect corresponds to the substantive conception of whether they can consciously access their reasons for doing so or not. This applies, for instance, to the traditional Filipino mindset that being in a relationship with a partner implies that one is the other's property, though this tends to work asymmetrically in favor of men. If John and Mary are in a relationship, then Mary has an obligation to have sex with John when he pleases. Payongayong, however, points out that attitudes are evolving amongst women, such as those who

accept the concept of open relationships or reject the concept of monogamous relationships. Such women believe that since they are not the property of men, they are free to refuse to have sex with their partners and can articulate their reasons for their refusal. Hence, they can be said to be sexually autonomous in both the procedural and substantive senses.

### Deepfake Pornography

We have now laid out two liberal feminist standards upon which the moral justification for criminalizing deepfake pornography under the Safe Spaces Act can rest. But before that, it will be helpful to distinguish the philosophical issues surrounding pornography *in general* from those of deepfake pornography in particular.

Ever since the 1970s, liberal feminists have raised various objections against the existence of pornography, some of which apply to deepfake pornography to varying degrees. Andrea Dworkin (1978, 217), for instance, referred to pornography as “the new terrorism” due to its sexual degradation of women through images of their being bound, bruised, and maimed. Catharine MacKinnon (1989, 329), on the other hand, argues that pornography creates only “the appearance of choice or consent.” There are two levels of harm created here: the first is that the actresses who participate in porn are made to appear to consent to the sexual positions being forced upon them by male actors and directors when in reality they often do not, and the second is that female sexuality becomes identified with “love of violation, variously termed female masochism and consent,” such that women in general become thought of by men as *enjoying* degrading acts being performed upon them (MacKinnon 1989, 329). A slightly different point is made by Rae Langton and Caroline West, who explain that women are “silenced” by pornography, which they compare to a Wittgensteinian language game governed by certain rules, many of which do not resemble those which underlie ordinary conversations. There are two kinds of silencing in play here. The first is on the level of the language game played by the pornographer and the actress. Quite literally, because the latter is being paid under contract to follow whatever the former directs her to do, she is silenced into following whatever the script tells her to do. She may, for instance, utter words of pleasure and pretend to orgasm even when she experiences pain and disgust. The second is on the level of the language game played by everyday men and women who consume pornography, especially scenes where women are scripted to say ‘*No!*’ to a man’s sexual advances but are instructed to evince pleasure. In other words, this kind of silencing involves utterances which are interpreted by audiences to mean the exact opposite of what the speaker intends to convey. Consequently, men are conditioned into thinking that ‘*No!*’ means ‘*Yes!*’, and can force their sexual partners into humiliating and degrading sex acts while thinking that they are actually enjoying it (Langton and West 1999, 313-314). Meanwhile, Martha Nussbaum argues that pornography violates the Kantian principle of treating each person as an end rather than a means, because the message that it sends male consumers is that women are just objects for sexual gratification, whereas men are the agents with subjectivity and autonomy. That is to say, in commodifying women through *Playboy* magazine or hardcore pornographic films, men are trained to think that women are no different from expensive cars, suits, or other status symbols in the

world of men—trophies to be collected rather than persons to be valued in and of themselves (2006, 283-284).<sup>3</sup> Furthermore, Helen Longino raises a different concern: pornography violates the right to privacy. In her view, while the right to privacy might protect the individual consumption of pornographic images or videos—such as when a woman voluntarily takes nude photos or videos of herself for the exclusive use of her partner—it does not necessarily protect their distribution to the general public, especially in cases lacking her consent. The parallel she draws concerns the crime of defamation. Whereas it is not criminal to harbor ill thoughts about other people, making these thoughts public through libel or slander is, because other people become involved (2017, 234-235). Finally, Mari Mikkola discusses what she refers to as the “personification” of pornography: men personify pornography when they use pornography to satisfy their sexual desires. The example she gives is that of using pornographic magazines to move from mere sexual arousal to real sexual release (ejaculation). It differs from using food magazines and other forms of representation. Looking at the picture of a hamburger can trigger hunger in someone, but the representation thereof can never satisfy that hunger. It is different in the case of sexual representations like pornographic magazines or videos—the consumption thereof not only triggers sexual desire, but it is also sufficient to satisfy it. And so the object itself must be a sex object; the consumption of magazines is to literally have sex with pieces of paper, while the consumption of videos is to literally have sex with pixels (2019, 254-255). The problem here is that personifying pornography places sexual objects and actual women in the same ontological category, which *dehumanizes* women (Vadas 2005, 188). Mikkola distinguishes between objectification and dehumanization. In her view, dehumanization involves derivatization, or the phenomenon wherein the subjectivity of someone is made wholly dependent on that of another. Quite literally, *x* is derived from *y* if the nature of *x* ontologically depends on that of *y*. In this case, women are reduced into the carnal desires of men, to whom pornography grants the power of conveniently switching between treating women as autonomous agents sometimes and non-autonomous objects in others through the personification of pornography. While Mikkola herself casts doubt on this conflation—and indeed she refers to it as a category mistake—there is nevertheless something novel about this view lacking in earlier literature about objectification.

There are several other arguments raised against pornography in general, but our brief survey highlights six that are of relevance to our paper. Pornography (1) is an act of violence that degrades women, (2) manufactures the appearance of consent, (3) silences and distorts their communicative intentions (4) objectifies and strips them of agency for the purpose of satisfying male pleasure, (5) violates their right to privacy, and (6) dehumanizes them. We now discuss how these harms apply to and may in fact be amplified in cases of deepfake pornography.

**(i) Deepfake Pornography and the Objectification of Women.** Like other forms of pornography, deepfake pornography degrades, dehumanizes, and silences women into sexual objects whose primary purpose is to satisfy the carnal pleasures of men. There are, however, new problems raised by artificial intelligence that exacerbate these harms. The obvious new harm that is presented is that now *anybody* can be victimized by deepfake pornography, not just women who have private photos or videos. The immediate target of deepfake pornography is women whom men fantasize

about, such as celebrities like Scarlett Johansson, Margot Robbie, and Taylor Swift, and locally, Angel Aquino, who described her video as “dehumanizing” and “nauseating” (Tamayo 2025). Another new set of vulnerable targets involves minors (especially popular young adult actors) who often do not even have sexual histories and are preyed upon by pedophiles and child molesters who generate videos of them in very sexually demeaning and mature positions (Pascale 2023, 353). Apart from being made to appear to engage in sex acts, women can be made to utter humiliating words and sounds that reduce them to submissive sex toys at the mercy of men. Worse, several deepfake videos are uploaded to mainstream websites like *Pornhub* where anybody can stream or download them for free, which means the audience before which a woman can be humiliated can range globally in the millions.

A further unique harm occurs when deepfake technology is used to superimpose a woman’s face onto another body, or when sensitive body parts like a woman’s breasts or genitals are digitally manipulated to make them more enticing. Apart from the fact that such alterations cause deep trauma, public humiliation, and economic exploitation (on websites that require payment before viewing), they also trigger body dysmorphia by making women feel that their natural bodies are inadequate without enhancement. This especially holds true in cases wherein the faces of ordinary women are superimposed on the bodies of pornographic actresses to whom they will inevitably be compared (Dilluvio 2024, 52-54; Hunt 2025, 21) Apart from these insecurities, deepfake videos also create a sense of paranoia among women who feel that men in public are constantly paying attention to specific body parts, and that they are reduced into breasts or genitals that are primarily valued for sex only. The reduction of women from human beings to body parts is a contemporary manifestation of the dehumanization Mikkola talked about: a woman’s self-identity, worth, self-image, and personhood can become inadvertently dependent upon the opinions and perceptions of men.

It might be objected that women are not really objectified by deepfake pornography because no actual women are involved in their production, unlike in traditional forms of pornography. The “women” involved, the objection goes, are only comprised of pixels and are, at most, mere representations of women. But, the objection concludes, representations of women are not actual women, and hence, there is no actual person being objectified. We believe that this objection is faulty, however, because it misunderstands the nature of representation. Even if they are representations, when they are consumed by humans, they are still objects of intentionality. By ‘intentionality’ we mean the feature of human consciousness that gives it the capacity to be *about* or *directed* at something. Searle (1983) argues, however, that intentionality extends to representations like photos and videos. When we look at a picture of our dog, for instance, our mental states are not just directed at a piece of photo paper. They are directed at the object which is *being* represented by the video or image—our actual dog that exists in the world. Similarly, the intentionality of consumers of deepfake pornography is not only directed at pixels precisely because these pixels are representations of something else—actual women. In fact, part of what makes deepfake pornography so titillating for consumers is the fact that their mental states are *about* that person, and deepfake pornography is created in a way to intentionally construct or represent a specific individual. It is no accident that representations of people in these videos look like Scarlett Johansson, because they are

intended to represent and are structurally directed by their creators at the actual Scarlett Johansson. The representation is merely an intermediate layer between the consumer and the actual object. But the intentionality, at the end of the day, is *still* directed at the actual object. Hence, *contra* the objection, deepfake pornography still constitutes an objectification of women through the nature of representation and intentionality, even if the immediate object of intentionality is not an actual living person.

**(b) Deepfake Pornography and the Vitiating of Sexual Autonomy.**

Deepfake pornography, like traditional pornography, vitiates the sexual autonomy of women by creating the illusion of consent, stripping them of agency, and violating their right to privacy to varying degrees. As we have seen above, its creators are capable of superimposing a woman's face onto another person's body and making it appear as if she consents to lewd and obscene sexual acts that are created digitally. There is a deeper level of harm here, however, and it is that unlike cases wherein professional actors agree to film pornographic movies or participate in nude photoshoots, or even cases in which authentic nude photos or private sex videos filmed between consenting couples are leaked online without their consent, cases of deepfake pornography may involve women who have not yet consented to any sexual act whatsoever, or at least the recording thereof. While there are admittedly women who find thrill in recording intimate videos with their partners, and women who consent to having sexual images of themselves shared publicly, there are several other women who have either never consented to any kind of recording, let alone women who have not yet even decided that they are ready to engage in sexual acts at all. This may be due to their preference for privacy, their religious beliefs with respect to sex, their choice of an asexual lifestyle, or their warranted fear that nude images may fall into the wrong hands. The obvious victims here are minors who are often at risk of being victimized by deepfake pornography (Furizal et al. 2025, 2) or young adults who might be the subject of blackmail and "revenge pornography" using deepfake technology, which is a phenomenon that is unfortunately on the rise in the Philippines (Blancaflor et al. 2024). Not only are they not given any choice concerning the distribution of sexual images in their likeness (Harris 2021, 13374), they may not even have chosen to participate in any sexual activity at all. The trauma and loss of sexual autonomy for parties who do not even view themselves as sexual or sufficiently mature individuals may therefore be even greater than in cases of non-deepfake pornography. To exacerbate the problem further, deepfake technology has advanced so rapidly that pornographic videos and images look increasingly authentic to the point where it can be difficult to distinguish between authentic and deepfake images. In fact, deepfake creators more often than not deliberately try to make their videos appear authentic in a number of ways and for a number of reasons. For instance, representing a woman as enthusiastically making faces or moaning in pleasure could be because authentic-looking videos are more sexually arousing, or because their attempts to blackmail their subjects are likely to be more successful, or because consumers are more willing to pay good money for videos that pass as realistic (Easterbrook-Smith 2025, 5). The harm, of course, is a total loss of agency and control not just over one's own body, but one's image and likeness. They are stripped of the choice not only as to who can see sensitive images of their intimate moments, but as to who can even see representations of them at all.

There are many other social, economic, and psychological harms that are caused by deepfake pornography, but they are beyond the scope of this paper. Now that we have established that there are moral justifications for liberal feminists to reject deepfake pornography, we now turn to the legal justification for including it within the scope of the Philippine Safe Spaces Act, particularly from a jurisprudential perspective that seeks to clarify the nature of law.

## THE EXPANSION OF THE SAFE SPACES ACT AND DEEPPFAKE PORNOGRAPHY

### Specifying Deepfake Pornography within the Philippine Safe Spaces Act

From the preceding discussion, there has been progress to widen the scope of the Philippine Safe Spaces Act (Republic Act. No 11313) to include online forms of sexual harassment brought about by the increased development and use of digital and AI technologies. Specifically, under Section 12 of the proposed Senate Bill amending Republic Act. No 11313 (S. No. 2897), it expands its scope by stating that “gender-based online sexual harassment includes acts that use information and communications technology, including *artificial intelligence and other emerging technologies*” (Senate of the Philippines 2025). Although this recognizes an extensive list of harmful online sexual acts committed within virtual spaces (such as cyberstalking and online grooming), deepfake pornography is currently only implied within this legal definition and thereby fails to encapsulate the full range and severity of the harms it entails. Thus, we argue that the explicit and unequivocal inclusion of deepfake pornography under the Safe Spaces Act as a legally recognized form of online sexual harassment is necessary for the legal protection of women against objectification and violation of sexual autonomy. This will also aid in recognizing tech-based abuse under Philippine law.

Although deepfake media is justified as a means of creative expression that does not merit any need for legal regulation, Furizal et al. (2025) state that deepfakes, within the context of pornography, often become a tool for exploitation, extortion, or even public humiliation. Beyond spreading disinformation, deepfake pornography poses distinct vulnerabilities that must be considered. Firstly, it enables the unsolicited use of an *actual* person’s image and likeness to create a harmful, fabricated narrative. Unlike other synthetic (AI-generated) forms of pornography, deepfake porn capitalizes on the use of real identities and actual content posted from social media. Secondly, the scope of threat extends to *anyone*; any unsuspecting person can be targeted and victimized, regardless of their status and age. Thirdly, the perpetrator’s anonymity poses difficulty on whom to hold *accountable*. Lastly, the *ease of its transmission and integration* to open-access sites exacerbates the intensity of the harms (e.g., invasion of privacy, harm to reputation, etc.) the subject may experience through increased exposure. Even if its harmful tendency of enabling objectification and compromising a person’s sexual autonomy overlaps with other acts of online sexual harassment, the extensive reach of deepfake porn and its unpredictability covertly threatens the safety of virtual and physical spaces that warrant the need for legal protection through the Safe Spaces Act — which aims to ensure that spaces are free from gender-based harassment. That said, the succeeding paragraphs shall present justifications rooted in

analytic legal philosophy to establish the connection between jurisprudence, liberal feminist principles on objectification and sexual autonomy, and Philippine law, which will support the claim for including deepfake pornography as a distinct act of online sexual harassment under the Philippine Safe Spaces Act.

## **Jurisprudence, Liberal Feminism, and the Safe Spaces Act**

### ***Special Jurisprudence and Liberal Feminist Principles***

In this case, we turn to examine how liberal feminist principles can figure into law through the lens of *special jurisprudence* — which deals with understanding more specific and granular law-dependent concepts as applied by the legal system (Green 2023, 128).<sup>4</sup> This is so because feminist legal theory has been concerned with identifying, addressing, and criticizing particular legal rules in a specific legal system that is thought to be patriarchal which puts women at a legal disadvantage both in theory and practice. Moreover, gender norms have the capacity to influence the *content and application* of law, especially to legal rules related to addressing gender-based matters (Green 2023, 148).

In defense of analytical jurisprudence, Green’s analysis on the legal concept of marriage exemplifies how feminist legal theory can influence and reshape marriage law.<sup>5</sup> Influential feminist social movements, such as the 1960s Women’s Liberation Movement, called for a radical social change emphasizing the rights of women. This eventually led to a conceptual change in marriage law due to the formal legislation of divorce. This has been possible because marriage law is a civil contract based on consent (*consensus*). As explicated by Cruz (2024), Liberal Feminist Jurisprudence can be applied as a framework of principles to support legal rules to be gender equitable. Specifically, she claimed that the consideration of liberal feminist principles can unmask patriarchal undertones in Philippine law and bring to light women’s narratives and experiences. Moreover, in application to secularizing Philippine divorce legislation, the possibility of conceiving a progressive concept of family and marriage in the Philippines to empower Filipina women becomes plausible. Since liberal feminist principles can influence and shape legal rules by means of social and conceptual change, we will now examine how liberal feminist principles can figure into the Philippine Safe Spaces Act to extend legal protection against deepfake pornography.

### ***Consideration of Liberal Feminist Principles in the Philippine Safe Spaces Act***

In legal philosophy, particularly espoused by Dworkin, legal rules and principles are thought to be logically distinct from each other as a significant criticism against legal positivism, as being incapable of considering legal principles. Where rules are absolute and noncontradictory, principles sustain an evaluative dimension in the practice of law (Dworkin 1977, 22-27). In other words, principles encapsulate and enforce moral standards on relevant legal rules, which eventually led to his interpretivist view of law — where the law “follows from the most morally attractive interpretation of a community’s legal materials” (Adams 2017, 622). However, Raz (1972, 832-833) dissolves the rule-principle distinction by clarifying that, much like principles, legal rules also bear weight, which explicates their tendency to conflict with other rules as well as relevant principles. Through this redefinition, Raz reasons that

principles can justify rules that allow it to become more flexible and broadly applicable. In elaboration, with principles serving as an indirect guide to human behavior, it has the weight and relevance to widen or narrow down the scope of a rule.

With this, we can justify the inclusion of liberal feminist principles grounded in gender norms and women's experiences to make the expansion of the Philippine Safe Spaces Act even more amenable to legal specifications that prioritize women's welfare and the safety of those who become subject to various acts of sexual harassment. This is also consistent with the idea that law, possessing *de facto authority* to guide human behavior, should provide "publicly ascertainable standards" that do not exclude any group (Raz 1994, 51).<sup>6</sup> Although Philippine legislation appeals to the moral sense of what is right or wrong, fair and just, a legal rule's validity is established on the grounds of formal declaration asserted by human action under the formal legislative process, not the degree to which the law conforms to moral correctness (Adams 2025, 564).<sup>7</sup>

With this, it is plausible to consider liberal feminist principles to justify certain amendments that make legal rules more reflective of the increasingly liberal gender norms emerging in the Philippines. In Razian terms, *principles* proscribe highly unspecified acts (such as upholding the right to equality), whereas legal rules proscribe highly specified acts (e.g., speeding and murder) (Raz 1972, 838). From this, we can infer that liberal feminist principles proscribe the objectification of women and condemn any violation that compromises a person's sexual autonomy, especially that of a woman. It is also worth noting that these principles do not necessarily contradict our moral intuitions, as they aim to uphold the right to equality and freedom.

It is also worth recognizing the reasoning and principles that justify the moral incorrectness of sexual harassment acts that bear weight on its prohibition in Philippine law, which align with the inclusion of deepfake pornography. Moreover, the legislative authority's discretion can cover these specific actions, considering the principles that justify the amendments under the relevant legal rule, thereby providing clear grounds for legal protection and punishment.

In other words, these liberal feminist principles cover a broad set of actions, norms, and social facts that can be used to justify the Philippine Safe Spaces Act, which has the force to prohibit specific acts of sexual harassment. Having said that, we will now analyze how two liberal feminist principles that were used to justify the rejection of deepfake pornography on moral grounds can figure into the Philippine Safe Spaces Act. For the sake of brevity, each liberal feminist principle cited will be understood as the *Principle of Rejecting Sexual Objectification (PRSO)* and the *Principle of Protecting Sexual Autonomy (PPSA)*.

(i) **The Principle of Rejecting Sexual Objectification (PRSO)** upholds the view that women should not be treated as objects for the sexual gratification of men. From the first part of this paper, through objectification, women are placed in subordination to men, which exacerbates the inequalities they face in legal, social, and cultural contexts. It also devalues their agency as it violates the Kantian principle of treating a person not as a means to an end (Nussbaum, 2006). With this, PRSO can provide justification for women's safety and the right to equal treatment. In this regard, this principle is compatible with supplementing the expanded Safe Spaces Act because it aligns with the specific act's aim of promoting safe virtual and physical spaces free from gender-based harassment and discrimination. In extension, advocating for legal protection against

victimization through deepfake pornography can be clearly expressed within the scope of the legal rule as it becomes acquiescent towards the consideration of current gender norms aligned with liberal feminist principles that aim to address pre-existing laws and practices that exclude or disadvantage women and other minority groups.

(ii) **The Principle of Protecting Sexual Autonomy (PPSA)**, on the other hand, is concerned with acknowledging women's capacity in determining how they engage in sexual activities (procedural sexual autonomy) and the corresponding reasons that justify their desires and preferences, which has a direct correlation to the way they exercise their sexual autonomy (substantive sexual autonomy). That said, this principle supports the view that women should not be silenced or manipulated, borne from acts that restrict and subvert their sexual autonomy. For this reason, PPSA, as an incorporated principle in the Safe Spaces Act, can honor women's choices and preferences in this matter. More concretely, though, it can support the specific act of prohibiting the creation and transmission of nonconsensual pornographic content, especially one produced through deepfake technology. Moreover, such acts can become explicitly punishable under the grounds of the Philippine Safe Spaces Act.

Through this, broad liberal feminist principles can provide justification for expanding the Safe Spaces Act and for the specific acts of sexual harassment and discrimination it aims to extinguish. Although Philippine law has pre-existing constitutional and legislative provisions that recognize the equality of women, notwithstanding its partiality for patriarchal principles, incorporating these principles plays a complementary and significant role in Philippine law to justify the specific legal protections needed to address the threats and vulnerabilities arising from rapid technological advancements.

### ***Legislating the Expanded Safe Spaces Act, including Deepfake Pornography***

Upon reviewing the status of the Senate Bill amending the scope of the Safe Spaces Act (S. No. 2897), it must be noted that it has already passed on third reading, with unanimous approval from the Senate. In the legislative process, bills under the third reading no longer entertain further amendments. Although amendments are possible through a conference between the bicameral committee after the third reading, their eligibility and urgency for further amendments seem unlikely.<sup>8</sup> This is due to other legislative priorities, such as bills that specify legal provisions on disaster resilience and economic reform, whose impact is deemed even more critical to aid the development of the Philippines and the welfare of its citizens. Despite this, we argue that incorporating specific legal provisions on deepfake pornography as a distinct act within the legal rule can criminalize it without adding a legislative burden related to the creation of a new law. Moreover, specifying deepfake pornography under the Safe Spaces Act is a step to further fortify legal protection against online sexual harassment driven by AI that has repercussions for the production, management, and control of digital data and its security in online spaces. In this way, deepfake pornography can be categorized as being part of a cluster of online acts of sexual harassment, while reducing the difficulty of identifying it that may be obscured under a broad legal definition.<sup>9</sup>

In connection with legal philosophy, the unanimous vote of approval regarding the expansion of the Safe Spaces Act is an indicator that the Philippine legislative body, despite possessing an inclination that favors patriarchal and dogmatic Catholic

principles, is receptive to adopting changes to improve and uphold women's safety, autonomy, and equality by means of law. Adams' Criteria of Validity (2025) emphasize the importance of lawmakers who consider relevant principles to anticipate legal vulnerabilities. Moreover, the authoritative capacity to subject valid laws into a process of invalidation when proven to be unconstitutional ensures that laws remain consistent and amenable to future legal developments. Moreover, Raz (1972) notes that when principles conflict, a weighing mechanism prioritizes those that refer to pre-existing norms, rules, and principles that are significantly more important and relevant, thereby making legal rules and policies equitable and flexible. In application, it is possible to conceive that the expanded Safe Spaces Act, including deepfake pornography, can be legally valid since it will undergo the rigorous legislative process for approval. It would also be constitutional since its provisions would be written in accordance with the statutes within the 1987 Philippine Constitution.<sup>10</sup>

In summary, the expanded Safe Spaces Act outlining specific legal provisions for addressing deepfake pornography can acknowledge its distinct harms, such as compromising a person's identity and their privacy. Secondly, incorporating liberal feminist principles into law gives it shape and force within the sphere of rapid technological advancement despite the pre-existing adherence to patriarchal principles within the Philippine legal system. Thirdly, it can provide a concrete legal definition and protection, which can ease the ambiguities present in identifying offensive acts that can be categorically classified under deepfake pornography.

### **Possible Areas of Development and Discussions Post-legislation**

Since it is legally justifiable to expand the Safe Spaces Act to include deepfake pornography as a distinct act of online sexual harassment, its legislation can prompt the development of authentication techniques to pinpoint perpetrators and arrest the harms and damages that may be incurred as quickly as possible. Also, it can prompt streamlining tracking technologies to effectively locate and hold the perpetrator accountable, subject to the penalties comprehensively outlined in the expanded Philippine Safe Spaces Act.

Also, it can be understood in relation with other significant landmark legislations such as (R.A.) 11930 or *The Anti-Online Sexual Abuse or Exploitation of Children and Anti-Child Sexual Abuse or Exploitation Materials Act*. This law aims to improve and fortify specific legal protection for children from online sexual abuse and exploitation that entails greater penalties (e.g., lifetime imprisonment). Another policy that is applicable in conjunction with the aim of specifying deepfake pornography within Philippine law is (R.A.) 10175, also known as the *Cybercrime Prevention Act of 2012*. Under Chapter II, Section 4, offensive acts such as computer-related identity theft and unsolicited data interference are punishable by law, which can be used as further justification to criminalize deepfake pornography due to the scope of its harms and the manner in which it is produced and transmitted. That said, with clear grounds for deepfake pornography under the Safe Spaces Act, forms of indirect and synthetic sexual abuse can be accounted for, especially for children and women who are vulnerable targets of sexual objectification.

Moreover, it can underscore the need for additional legal protections for personal digital data and content across public networks. That said, there are other areas of concern that are subject to discussion within the area of emerging technologies; however, this paper emphasizes the prevalence of deepfake pornography and its severe threats against a person's sexual autonomy that warrant specific legal provisions under the Philippine Safe Spaces Act. With this, the succeeding discussion will explore the philosophical implications to further guide the inclusive legislation against deepfake pornography grounded in A.I. Ethics, specifying its place within an algorithm-centric digital space. More importantly on the side of AI technology, guiding principles on machine learning should also be expounded on.

## GUIDELINES IN LEGISLATION AND AI USE

The *Principle of Rejecting Sexual Objectification* (PRSO) and the *Principle of Protecting Sexual Autonomy* (PPSA) present an opportunity to guide legislation that strengthens regulation against the use of deepfake pornography. Even more important, such principles should be taken into consideration in machine learning to be able to guide AI decision-making itself. Below, we present several guidelines following these principles.

The first two guidelines have already been hinted at in the previous sections as they pertain to changes in legislation. Firstly, as previously established, both the PRSO and the PPSA emphasize the protection of consent, as outlined by Valentiner (2021). Therefore, clarification and redefinition of 'consent' in the context of AI-related legislation are imperative to protect PRSO and PPSA. The problem lies in the fact that, in AI deepfakes, consent is non-existent. The act of uploading a photo of oneself or on another's behalf does not in any way constitute consent for it to be manipulated by others without the subject's express permission, but deepfake pornography that is generated through AI may deceptively make it seem so, because it is the very nature of deepfake pornography to present a manipulated video as an authentic one. The laws should reflect such nuance, especially in the case of a patriarchal system that may often lead to outright blaming of the woman. This previous point leads us to our second guideline. The legislature should take steps toward expanding and specifying the terms in which "deepfake" may be used as a violation of such principles. This will further establish just how uncontrollable the situation is for the offended woman, considering a valuable difference between deepfake and other forms of online harassment. As of now, deepfake pornography is only implied in these laws, which is insufficient given the gravity of the offense previously mentioned. The said definition is already outside the scope of this paper, but the grounds for why it should be considered more closely have already been tackled. The guiding principles should extend towards related laws not limited to R.A. 11313. The Anti-Photo and Video Voyeurism Act (R.A. 9995), The Anti-Child Pornography Act, and even the pending legislation of the E-VAWC, among others, would all benefit from redefining the legal meaning and concept of consent.

Thirdly, and perhaps most importantly, a philosophical analysis of the issue need not be limited to a redirection of guiding principles in legislation. In fact, the very tool that is used, generative AI, may guide itself. Machine learning propagates the training

data it is fed, creating an algorithmic bias (Danks and London 2017). Not all algorithmic bias is inherently unethical (Silberg and Manyika 2019) but may be when there is an underrepresentation or overrepresentation of a data set pertaining to a specific group over another. The problem that arises in this case of upholding PRSO and PPSA is that AI, through algorithmic bias, may further worsen the problem of gender oppression online. An analogous situation can be drawn from the existing use of AI in the workplace and in job recruitment. Since training data dictates algorithmic bias, an overly skewed data set from a specific group of people may favor, among others, a particular race, age, or sex over others. For example, in assessing the credentials of a biological man and a woman and making the decision on who to hire, skewed information collected in favor of a man becomes discriminatory when AI misrepresents the irrelevant data (such as height, physical activity, and other differences which may be affected by sex) to be deemed as relevant in the decision of hiring (Ajunwa and Schlund, 2020). This is the case of “over” or “under” representation caused by skewed datasets. Essentially, we posit that in the context of deepfake pornography, there can even be a “false representation” of a population or certain groups of women that are affected by these deepfake materials which cause a discriminatory decision to be made against them. This has far-reaching consequences, as in the case of the aforementioned employment discrimination. For instance, in a job application that harnesses AI decision-making, background checks conducted by such AI may inadvertently take deepfake data into account. The false representation that a woman was seen in a deepfake video, without proper discernment by an AI decision-making scheme, can lead to termination by an employer and, even worse, discrimination in all other aspects, including background checks, when she applies for other jobs. The specific action step to be taken should then be an additional human intervention that guides the identification of deepfakes and proper discernment. A human-in-the-loop decision-making (Enarsson et al., 2022) scheme is logical because it can lead to further probing into the actual victim’s background, which would result in a non-discriminatory decision. That is, in the case where an AI algorithm cannot tell the difference between a deepfake and a real video, humans have a duty to carefully scrutinize and assess the facts in line with other credentials and determine whether such material holds any weight or even relevance in the final decision. It is this recognition that the women in said videos are victims and should not in any way be held accountable through a discriminatory decision. Of course, another step is to train the AI to better distinguish a deepfake video from a real video. The specifics of which are outside of the scope of this paper, but it seems that there is no guarantee as of now. Perhaps the case may be that technology and constant evolution in machine learning will lead to a more advanced discernment than that of human-in-the-loop, but as of now this is speculative.

Fourth, in favor of victims of deepfake pornography, the power to mark content as AI-generated to warrant its automatic deletion from the internet should be further explored. This comes in two ways. First is in the generation of such material, there should be a requirement that it must showcase that such material was AI-generated, such as through a watermark, and if potentially detrimental to the rights of a woman, should not even be generated. Secondly, in case of post-generated editing that removes marked warning, further discernment of deepfake by the same AI tool must be explored further. There should be steps taken in developing a technology that can ascertain a material’s veracity. It is like the “fact-checking” we see in social media, but

this measure is much more stringent and efficacious in that there would be automatic deletion. These effects of deepfake content have already been explored in the previous sections, and as such we believe that even the simple act of posting the material, even though it is obviously deepfake, has everlasting detrimental effects on a woman.

In sum, we posit that the governing guidelines are to enact legal amendments and new policies that reflect the new terms in relation to deepfake pornography, to provide safeguards against algorithmic bias which may integrate incriminating deepfake pornographic content, and to explore the power to streamline accessible methods for automatic deletion that can be effectively implemented.

## THE RIGHT TO FREE SPEECH

The strongest principled argument against the regulation of deepfake pornography is that it is a protected form of free speech. Indeed, in *Ashcroft v. Free Speech Coalition* (2002), the U.S. Supreme Court struck down parts of a federal law that tried to prohibit the production of virtual child pornography. The rationale of the court was that while the production of child pornography using real children remains a crime because it creates actual victims, the production of *virtual* child pornography does not create any victims, and thus the federal law infringed on protected speech. Deepfake pornography is a form of virtual pornography and is thus entitled to the same protection following the court's rationale. For those who espouse this view, their general argument is that it is part and parcel of a liberal democratic society that people use free speech for sometimes undesirable purposes. While the creation and consumption of deepfake pornography is morally unfortunate, there is a greater value in protecting a marketplace of ideas, which is the discovery of truth (Franks & Waldman 2019; Gieseke 2020). They then compare deepfake to forms of art, political commentary, or satire, none of which are "true" in the literal sense, and yet are protected as forms of free speech.

We raise two objections against this reasoning because we believe that the principle of free speech—even as conceived by proponents of this argument—has been misapplied. The first is in the claim that deepfake pornography, like other forms of art, serves the pursuit of truth. On the contrary, we argue that deepfake pornography aims to distort the truth, not reveal it. Art and satire use abstract representations, metaphor, and hyperbole to emphasize deeper messages. Often, these messages may be political or moral in nature. Deepfake pornography, however, is merely created for lewd entertainment and does not help uncover any valuable political or moral truth. Moreover, deepfake pornography by its very name and nature creates *false* representations of individuals. These are different from creating *exaggerated* representations of individuals, which deliberately exaggerate their elements to elicit humor, anger, or awareness of a message they aim to bring to light. That is, their creators *hope* that their audiences understand the exaggerated elements of their work to bring their attention to the deeper message. But deepfake pornography aims to do the opposite: which is to pass off as authentic the images and videos of its victims, precisely because the purpose of deepfake pornography is to titillate consumers with its realism and make them believe that they are consuming *actual* pictures and videos of their victims.

The second objection is that deepfake pornography *does* create victims. They are the actual women who are presented as engaging in demeaning sexual acts against their consent. The argument, which is reinforced by the unfortunate decision to *Ashcroft*, seems to rest on a very narrow conception of harm that is based on physical and psychological harm that is accrued in the physical production of pornography. As we have argued, however, harm has several dimensions beyond the physical. Victims suffer psychological, emotional, social, and even financial harms from the production of deepfake pornography, and as courts have banned acts of free speech that have led to both physical and non-physical forms of harm (e.g., incitement), we believe that they should likewise prohibit deepfake pornography on similar grounds. We conclude that free speech cannot be used to protect deepfake pornography as presented.

## CONCLUSION

In Part I, we introduced the framework of liberal feminism and two of its goals in relation to pornography that provide the moral justification for expanding the Philippine Safe Spaces Act to include deepfake pornography within its ambit: (1) the rejection of objectification and (2) the protection of sexual autonomy. In Part II, we formulated *PRSO* and *PPSA* as legal principles that cover a broad set of actions and norms that, in turn, provide the jurisprudential justification for expanding the Philippine Safe Spaces Act. In Part III, we applied our insights from Parts I and II to formulate recommendations rooted in AI ethics from a feminist standpoint for more concrete amendments to relevant laws and AI-policy guidelines that aim to protect women. In Part IV, we responded to the objection that the regulation of deepfake pornography amounts to an assault on free speech. If the arguments we have presented in this paper are sound, then we hope we have been able to provide a philosophical foundation for one of our most important current pieces of legislation.

## NOTES

1. In September 1997, the Philippine Congress passed Republic Act No. 8353 (the “Anti-Rape Law”). Unfortunately, specific clauses within that bill have generated controversy, such as the following clause which decriminalizes marital rape: “Article 266-C. Effect of Pardon. The subsequent valid marriage between the offended party shall extinguish the criminal action or the penalty imposed. In case it is the legal husband who is the offender, the subsequent forgiveness by the wife as the offended party shall extinguish the criminal action or the penalty: Provided, that the crime shall not be distinguished or the penalty shall not be abated if the marriage is void *ab initio*.”

2. Valentiner does not use the terms ‘procedural’ and ‘substantive’, but she offers a model of substantive sexual autonomy that depends on five conditions: “Consent is achieved when (1) actors who are capable of communicating and understanding (2) meet at eye level; (3) reach an agreement on sexual action through negotiation while maintaining each person’s autonomy; (4) while acting in accordance with this agreement; (5) and ensuring continuity of consent in terms of time.” (p. 706)

3. Nussbaum's view here is part of her broader theory of objectification which consists of seven notions: (1) Instrumentality: the objectifier treats the object as a tool of his or her purposes; (2) Denial of autonomy: The objectifier treats the object as lacking in autonomy and self determination; (3) Inertness: The objectifier treats the object as lacking in agency, and perhaps also in activity; (4) Fungibility: The objectifier treats the object as interchangeable (a) with other objects of the same type, and/or (b) with objects of other types; (5) Violability: The objectifier treats the object as lacking in boundary-integrity, as something that it is permissible to break up, smash, break into; (6) Ownership: The objectifier treats the object as something that is owned by another, can be bought, or sold, etc.; (7) Denial of subjectivity: The objectifier treats the object as something whose experience and feelings (if any) need not be taken into account. (2006, 257)

4. In contrast to special jurisprudence, general jurisprudence is concerned with exploring questions related to the conceptual nature of law such as identifying the structure and features of legal systems *in general* (Green 2023, 127).

5. In the *Germ of Justice*, Green addresses Conaghan's claim that analytical jurisprudence isolates legal concepts from gendered contexts within the realm of specific legal systems which necessitates the integration between feminist and general jurisprudence. However, Green points out that general jurisprudence is "well-equipped to detect the conceptual role of gender where it exists." Please see pp. 136 – 148 to get a detailed view of the debate.

6. Under Raz's Service Conception of Authority, *de facto* authority is when something in itself is recognized to be sufficient reason to justify whether something is permissible or impermissible (Raz 1994, 215).

7. In his Criteria of Validity, Adams specifies that from Dworkin's pedigree thesis, the manner in which laws are adopted or developed figure well into the criteria of validity as possessing a formal nature even if he does not concede with the idea of morality as having a direct substantive role in law (Adams 2025, 568).

8. Within the Philippine legal system, a *bicameral conference committee* is made for reconciling conflicting provisions within a proposed bill through a meeting between a commensurate number of representatives from the lower and higher houses of the Philippine Congress.

9. The term 'cluster' is used here to signify deepfake pornography as a distinct act categorized under the same offense, rather than being lumped into a legal definition where it is only implied.

10. Within the 1987 Philippine Constitution, there are many relevant statutes that recognize and protect women. These include Article II, Sec. 14, where the fundamental and legal equality of women is recognized, as well as Article X, Sec. 10, outlining the necessary legal provision for safe working conditions. Moreover, the Bill of Rights, Sec 1, states that women, just as much as men, are subject to the equal protection of laws.

## REFERENCES

Adams, Thomas. 2024. "Criteria of Validity." *The Modern Law Review* 88 (3): 557–84. <https://doi.org/10.1111/1468-2230.12938>.

- Adams, Thomas. 2017. "Law's Umpire." *Jurisprudence* 8 (3): 621–30. <https://doi.org/10.1080/20403313.2017.1385291>.
- Philippines. 2012. Republic Act No. 10175 (*Cybercrime Prevention Act of 2012*). September 12, 2012. [https://lawphil.net/statutes/repacts/ra2012/ra\\_10175\\_2012.html](https://lawphil.net/statutes/repacts/ra2012/ra_10175_2012.html).
- Philippines. 2019. Republic Act No. 11313 (*Safe Spaces Act*). April 17, 2019. [https://lawphil.net/statutes/repacts/ra2019/ra\\_11313\\_2019.html](https://lawphil.net/statutes/repacts/ra2019/ra_11313_2019.html).
- Philippines. 2022. Republic Act No. 11930 (*Anti-OSAEC and Anti-CSAEM Act*). July 30, 2022. [https://lawphil.net/statutes/repacts/ra2022/ra\\_11930\\_2022.html](https://lawphil.net/statutes/repacts/ra2022/ra_11930_2022.html).
- Aguilar, Delia D. 1989. "The Social Construction of the Filipino Woman." *International Journal of Intercultural Relations* 13 (4): 527-551.
- Ashcroft v. Free Speech Coalition. 535 U.S. 234 (2002).
- Barber, Alex. 2023. "Freedom of Expression Meets Deepfakes." *Synthese* 202 (40): 1-17.
- Benonguil, John Achilleus M. 2022. "Sexual Objectification of Women in Alcohol Print Advertisements in the Philippines." *American Journal of Interdisciplinary Research and Innovation* 1 (2): 67-70.
- Blacaflor, Eric, et al. 2024. "Deepfake Blackmailing on the Rise: The Burgeoning Posterity of Revenge Pornography in the Philippines." In *ICIIT '24: Proceedings of the 2024 9th International Conference on Intelligent Information Technology*, 295–301. <https://doi.org/10.1145/3654522.365454>.
- Bourgault, Jacob. 2025. "Free Speech and Synthetic Lies: Deepfakes, Synthetic Media, and the First Amendment." *Student Journal of Information Privacy Law* 3 (1): 50-73.
- Brooks, Tina, Princess G., Jesse Heatley, Jeremy J., Scott Kim, Samantha M., Sara Parks, et al. 2023. *Increasing Threat of Deepfake Identities*. Department of Homeland Security. [https://www.dhs.gov/sites/default/files/publications/increasing\\_threats\\_of\\_deepfake\\_identities\\_0.pdf](https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf).
- Cruz, Sophia. 2024. "The Filipina's Right to Un-marry: Liberal Feminist Jurisprudence and Divorce Legislation in the Philippines." Master's thesis, University of the Philippines Diliman.
- Dilluvio, Parera. 2024. "Deepfake: The Crossover Between Pornography and Artificial Intelligence." *Revista De Filosofie Aplicata* 7 (12-13): 46-59.
- Danks, David, and Alex John London. 2017. "Algorithmic Bias in Autonomous Systems." In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, edited by Carles Sierra, 4691–97. <https://doi.org/10.24963/ijcai.2017/654>.
- Dworkin, Andrea. 1978. "Pornography: The New Terrorism." *N.Y.U. Review of Law and Social Change* 8 (2): 215-218.
- Dworkin, Ronald. 1977. "The Model of Rules I." In *Taking Rights Seriously*, 14–45. Harvard University Press.
- Easterbrook-Smith, Gwynn. 2025. "Pornographic Aura, AI, and the Value of Authenticity." *Porn Studies* (June): 1-13.
- Enarsson, Therese, Lena Enqvist, and Markus Naarttijärvi. 2022. "Approaching the human in the loop—legal perspectives on hybrid human/algorithmic decision-making in three contexts." *Information & Communications Technology Law* 31 (1): 123-153.

- Franks, Mary Anne and Ari Ezra Waldman. 2019. "Sex, Lies, and Videotape: Deepfakes and Free Speech Delusions." *Maryland Law Review* 78 (2): 892-898.
- Furizal, Alfian Ma'arif. 2025. "Social, Legal, and Ethical Implications of AI-Generated Deepfake Pornography on Digital Platforms: A Systematic Literature Overview." *Social Sciences & Humanities Open* 12 (August): 1-25.
- Gieseke, Anne Pechenik. 2020. "The New Weapon of Choice: Law's Current Inability to Properly Address Deepfake Pornography." *Vanderbilt Law Review* 73 (5): 1479-1516.
- Gonzales, Christian. 2022. "Filling in the Loopholes: An Assessment of the Philippine Safe Spaces Act." *Res Militaries: European Journal of Military Studies*. 12 (5): 1177-1187.
- Green, Leslie. 2023. "The Germ of Justice." In *The Germ of Justice: Essays in General Jurisprudence*, edited by Leslie Green. Oxford University Press. <https://doi.org/10.1093/oso/9780192886941.003.0007>.
- Harris, Keith Raymond. 2021. "Video on Demand: What Deepfakes Do and How They Harm." *Synthese* 199 (September): 13373-13391.
- Hunt, Eva. 2025. "Deep-Fake Revenge Porn: The Intersection of Artificial Intelligence and Revenge Porn for a Modern Attack on Women." Master's thesis, California State University, Monterey Bay. [https://digitalcommons.csUMB.edu/caps\\_thes\\_all/2000](https://digitalcommons.csUMB.edu/caps_thes_all/2000).
- Langton, Rae and Caroline West. 1999. "Scorekeeping in a Pornographic Language Game." *Australasian Journal of Philosophy* 77 (3): 303-319.
- Longino, Helen. 2017. "Pornography, Oppression, and Freedom." In *Exploring Moral Problems: An Introductory Anthology*, edited by Steven M. Cahn and Andrew T. Forcehimes, 227-43. Oxford University Press.
- MacKinnon, Catherine A. 1989. "Sexuality, Pornography, and Method: Pleasure Under Patriarchy." *Ethics* 99 (2): 314-346.
- Mangalus, Jean. 2024. "Can the Philippines Confront Threats of Deepfake Porn?" *The Philippine Star*, September 25, 2024. <https://www.philstar.com/headlines/2024/09/25/2387697/can-philippines-confront-threats-deepfake-porn>.
- Martin, Jennifer L. and Julia Smith. 2020. "Why We March! Feminist Activism in Critical Times: Lessons from the Women's March on Washington." *Women's Studies International Forum* 81 (July): 1-11.
- Mikkola, Mari. 2019. *Pornography: A Philosophical Introduction*. New York: Oxford University Press.
- Miller, Shaun. 2022. "Sexual Autonomy and Sexual Consent." In *The Palgrave Handbook of Sexual Ethics*, edited by David Boonin, 247-270. Switzerland: Palgrave MacMillan.
- Nussbaum, Martha. 2006. "Objectification." *Philosophy and Public Affairs* 24 (4): 249-291.
- Pascale, Emily. 2023. "Deeply Dehumanizing, Degrading, and Violating: Deepfake Pornography and the Path to Legal Recourse." *Syracuse Law Review* (73): 335-366.
- Payongayong, Ma. Theresa. 2020. "Reflexive-Liberative Filipino Feminist Ethics." *Kritike* 14 (1): 64-97.

- Raz, Joseph. 1972. "Legal Principles and the Limits of Law." In Columbia Law School, Scholarship Archive. [https://scholarship.law.columbia.edu/cgi/viewcontent.cgi?article=1531&context=faculty\\_scholarship](https://scholarship.law.columbia.edu/cgi/viewcontent.cgi?article=1531&context=faculty_scholarship)
- Raz, Joseph. 1994. "Authority, Law, and Morality." In *Ethics in the Public Domain: Essays in the Morality of Law and Politics*, 210–37. Oxford University Press.
- Reyes-Espiritu, Ma. Adeinev. 2023. "Gender Matters in the Family: Confronting Gender Polarity in Philippine Transnational Families of Labour Migrants." *Journal of the European Society of Women in Theological Research* (31): 77-97.
- Russell Brenda et al. 2024. "What Makes a Liberal Feminist? Identifying Predictors of Heterosexual Women and Men's Liberal Feminist Ideology." *Analyses of Social Issues and Public Policy* 24 (1): 241-260.
- S.B. No. 2897, 19th Congress of the Republic, Third Regular Session (Phil.). <https://web.senate.gov.ph/lisdata/4547841599!.pdf>
- Searle, John. 1983. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
- Senate of the Philippines. 2025. "Bill Seeking to Expand Gender-Based Harassment to Include AI, Emerging Technologies Okayed." Press release, January 14, 2025. [https://web.senate.gov.ph/press\\_release/2025/0114\\_prib1.asp](https://web.senate.gov.ph/press_release/2025/0114_prib1.asp).
- Silberg, Jake, and James Manyika. 2019. "Notes from the AI frontier: Tackling bias in AI (and in humans)." *McKinsey Global Institute* 1 (6): 1-31.
- Tamayo, Bernadette E. 2025. "Actress Angel Aquino, Victim of 'Deepfake', Seeks Prosecution of Perpetrators of Cyber Pornography." *The Manila Times*, September 4, 2025. <https://www.manilatimes.net/2025/09/04/news/actress-angel-aquino-victim-of-deepfake-seeks-prosecution-of-perpetrators-of-cyber-pornography/2178939>.
- Tong, Rosemarie and Tina Fernandes Botts. 2018. *Feminist Thought*. 5th ed. New York: Routledge.
- Trott, Verity. 2020. "Networked Feminism: Counterpublics and the Intersectional Issues of #MeToo." *Feminist Media Studies* 21 (7): 1125-1142.
- Vadas, Melinda. 2005. "The Manufacture-for-Use of Pornography and Women's Inequality." *Journal of Political Philosophy* 13 (2): 174-193.
- Valdez, Isabel Kristine M. et al. 2022. "Violence Against Women in the Philippines: Barriers to Seeking Support." *The Lancet Regional Health-Western Pacific* 23 (May): 1-2.
- Valentiner, Dana-Sophia. 2021. "The Human Right to Sexual Autonomy." *German Law Journal* 22 (June): 703-717.

## ACKNOWLEDGEMENTS

The authors would like to acknowledge Professor Ma. Theresa T. Payongayong, PhD, and Sophia Beatriz G. Cruz for their valuable insights that contributed to the development of this article.