

The Criminalization Of Non-Consensual Deepfake Acts From The Perspective Of Positive Criminal Law And Human Rights Protection

Anggada Perkasa

Center of excellence for Criminal Law and Green Economy, Universitas Prima Indonesia, Medan 20118,
Indonesia

anggadaperkasa@unprimdn.ac.id

ABSTRACT

The development of generative artificial intelligence technology has given rise to the phenomenon of deepfakes, which involve the manipulation of a person's face, voice, or body language into synthetic audiovisual content that resembles the original. When this technology is used without the consent of the manipulated subject, particularly in the form of sexually suggestive content, fraud, or defamation, serious legal issues arise because Indonesian positive criminal law does not yet have norms that specifically and comprehensively regulate such acts. This study aims to analyze the position of non-consensual deepfakes within the framework of Indonesian positive criminal law and examine the urgency of their criminalization from the perspective of human rights protection. This study uses a normative juridical method with a legislative approach, a conceptual approach, and a case approach. The research findings indicate that relevant provisions are partially covered in the Electronic Information and Transactions Law, the Pornography Law, the Personal Data Protection Law, the Sexual Violence Law, and Law Number 1 of 2023 concerning the Criminal Code, which came into effect on January 2, 2026, specifically Articles 172 and 407 concerning pornography and Article 492 concerning fraud. However, these provisions do not explicitly address the elements of non-consent and the synthetic nature of a person's image as separate locus delicti, so enforcement still relies on general articles not designed for AI-based crimes. From a human rights perspective, non-consensual deepfakes constitute a violation of the right to privacy, the right to honor and dignity, and have the potential to constitute a form of online gender-based violence. This research recommends the need for a specific criminal norm (*lex specialis*) that explicitly regulates non-consensual deepfakes, along with strengthened redress mechanisms for victims.

Keywords: deepfake; non-consensual; criminalization; criminal law; human rights

INTRODUCTION

The advancement of generative artificial intelligence technology over the past decade has fundamentally changed the landscape of digital content production. One of its most controversial derivatives is deepfake, an image synthesis technique that utilizes generative adversarial network (GAN) algorithms and diffusion models to superimpose a person's face, voice, or gestures into videos, photos, or audio that are visually and auditorily indistinguishable

from the original recording. This technology was initially developed for positive purposes such as the creative industry, entertainment, and education, but in practice it is often misused to produce content that is detrimental to others, especially when done without the consent (consensus) of the person whose face or voice is manipulated.

The phenomenon of non-consensual deepfakes in Indonesia is no longer a hypothetical issue. From late 2025 to early 2026, the public was shocked by a series of cases of misuse of artificial intelligence-based image creation features on several global social media platforms, generating immoral content from the faces of public figures without their permission. These cases prompted a temporary suspension of access by the Ministry of Communication and Digital and sparked academic debate about the adequacy of positive criminal law instruments to address this type of AI-based crime. This phenomenon demonstrates that victims of non-consensual deepfakes, the majority of whom are women, experience multidimensional harm: psychological harm in the form of anxiety and trauma, social harm in the form of stigma and exclusion, and economic harm in the form of lost job opportunities and professional reputation.

Normatively, Indonesian criminal law currently regulates acts related to deepfakes partially through a number of sectoral laws, including the Electronic Information and Transactions Law (ITE Law) and its amendments, the Pornography Law, the Personal Data Protection Law (PDP Law), the Sexual Violence Crime Law (TPKS Law), and most recently, Law Number 1 of 2023 concerning the Criminal Code (New Criminal Code), which effectively came into effect on January 2, 2026. The problem is that these provisions are essentially designed to cover acts of pornography, fraud, defamation, or misuse of personal data in general, and do not explicitly consider the distinctive characteristics of deepfakes, namely the synthetic manipulation of a person's visual or vocal identity without their consent. As a result, law enforcement must construct laws analogously or extensively based on existing articles, which has the potential to create legal uncertainty and gaps in protection for victims.

From a human rights perspective, this issue cannot be separated from the constitutional guarantee of the right to privacy as stipulated in Article 28G paragraph (1) of the 1945 Constitution of the Republic of Indonesia, as well as international legal instruments such as Article 12 of the Universal Declaration of Human Rights (UDHR) and Article 17 of the International Covenant on Civil and Political Rights (ICCPR) which have been ratified by Indonesia through Law Number 12 of 2005. Non-consensual deepfakes, especially those with sexual content, are not merely a violation of moral norms, but also an attack on an individual's

control over their own image (bodily and image autonomy) as well as their honor and dignity as human beings. It is in this context that a study of the criminalization of non-consensual deepfakes becomes relevant to be carried out in depth.

Based on the background description, this research is formulated into two main questions. First, what is the status and adequacy of Indonesian positive criminal law norms in addressing non-consensual deepfake acts? Second, how urgent is the specific criminalization of non-consensual deepfakes from a human rights perspective? This research aims to provide a comprehensive legal mapping of the applicable legal framework and formulate academic arguments regarding the need for the establishment of specific criminal norms, as a contribution to the development of criminal law policies (penal policies) that are responsive to developments in artificial intelligence technology.

RESEARCH METHOD

This research is a normative legal research that positions law as a system of written norms analyzed doctrinally. The approach used includes three elements. First, the statute approach, which is used to examine and synchronize various relevant laws and regulations, namely the ITE Law and its amendments, the Pornography Law, the PDP Law, the TPKS Law, and the New Criminal Code. Second, the conceptual approach, which is used to elaborate legal concepts such as criminalization, consent, privacy, and honor in relation to the deepfake phenomenon. Third, the case approach, which is used to a limited extent to analyze cases of misuse of generative artificial intelligence features that occurred in Indonesia throughout 2025-2026 as an empirical illustration of the gaps in the norms studied.

The legal materials used consist of three types. Primary legal materials include binding laws and regulations, namely the 1945 Constitution of the Republic of Indonesia, the New Criminal Code (Law No. 1 of 2023), the ITE Law (Law No. 11 of 2008 as amended by Law No. 19 of 2016 and Law No. 1 of 2024), the Pornography Law (Law No. 44 of 2008), the PDP Law (Law No. 27 of 2022), and the TPKS Law (Law No. 12 of 2022), as well as relevant international legal instruments such as the UDHR and the ICCPR. Secondary legal materials include literature, scientific journals, studies by legal aid institutions, and legal opinions of practitioners discussing deepfakes and their criminal liability. Tertiary legal materials include legal dictionaries and encyclopedias to clarify technical terminology. The technique for collecting legal materials is carried out through library research, while the analysis is carried out

descriptively-qualitatively using a prescriptive method, namely not only explaining the applicable norms (das Sein) but also formulating the norms that should be (das Sollen) as policy recommendations.

RESULTS AND DISCUSSION

Characteristics of Non-Consensual Deepfakes as a Cybercrime Mode

Deepfakes are technically the product of deep learning algorithms that are able to reconstruct a person's facial features, expressions, voice intonation, and movement patterns based on a collection of available visual data, then combine them into new audiovisual content that is difficult to distinguish from the naked eye. Non-consensual elements arise when the engineering process is carried out without the consent of the subject whose data is used, giving rise to three main categories of *modus operandi*: non-consensual deepfake pornography, deepfakes for financial fraud and social engineering fraud, and deepfakes for defamation or political disinformation.

These three methods share the distinct characteristics of cybercrime: they are transnational in nature, can be replicated and disseminated massively in a short time through digital platforms, and are difficult to trace due to the perpetrators' ability to exploit online anonymity. The resulting harm is multi-layered, encompassing immaterial harm in the form of psychological suffering and privacy violations, as well as material harm in the form of lost income or job opportunities due to the damage to the victim's reputation. These characteristics distinguish non-consensual deepfakes from conventional forms of pornography and ordinary defamation, as the *locus of delicti* lies not solely in the distribution of content, but in the process of synthetically engineering a person's identity without their consent.

Indonesia's Positive Criminal Law Framework for Non-Consensual Deepfakes

Indonesian positive criminal law currently does not have a norm that explicitly mentions the term "deepfake," so law enforcement against such acts relies on extensive interpretation of a number of sectoral provisions. First, the ITE Law regulates the prohibition of the distribution of electronic information containing moral content through Article 27 paragraph (1) in conjunction with Article 45. In addition, Law Number 1 of 2024 as the second amendment to the ITE Law adds Article 35 which prohibits anyone from manipulating, creating, or changing

electronic information or documents with the intention of making the information appear authentic. This provision is conceptually closest to the character of deepfake because it targets acts of electronic data manipulation, although its formulation is still oriented towards the interests of document authentication and does not specifically target the engineering of a person's biometric image.

Second, the Pornography Law prohibits any act of producing, creating, reproducing, duplicating, distributing, or providing pornographic content. This provision is relevant when the resulting deepfake is sexually explicit, but the Pornography Law is not formulated to address situations where the victim's face is synthetically superimposed onto another person's body without their consent, so the lack of consent element is not a specifically defined element of the crime.

Third, the New Criminal Code (Law Number 1 of 2023) which came into effect on January 2, 2026 strengthens the provisions regarding pornography through Article 172 which formulates a broad definition of pornography to include images, animations, and other visual content containing obscenity or sexual exploitation, including those produced through digital technology. Article 407 of the New Criminal Code further threatens anyone who produces, makes, reproduces, duplicates, distributes, broadcasts, imports, exports, offers, sells, rents, or provides pornography with a minimum prison sentence of 6 (six) months and a maximum of 10 (ten) years or a fine of at least category IV, namely Rp200,000,000.00 (two hundred million rupiah) and a maximum of category VI, namely Rp2,000,000,000.00 (two billion rupiah). It should be noted that the enactment of the New Criminal Code also revokes Article 27 paragraph (1) in conjunction with Article 45 paragraph (1) of the ITE Law (formulation before the amendment to Law 1/2024), so that after January 2, 2026, the criminalization regime for electronic pornography will shift entirely to relying on the New Criminal Code. In addition, if deepfakes are used for fraudulent purposes, for example, engineering the voice or face of an official or company leader to induce the victim to hand over money, Article 492 of the New Criminal Code concerning the crime of fraud by trickery can be applied, with a maximum prison sentence of 4 (four) years.

Fourth, the PDP Law defines biometric data, including facial images and voice recordings, as specific personal data that requires explicit consent from the owner for any form of processing. Using a person's facial or voice image to create a deepfake without consent could potentially qualify as unlawful processing of personal data as regulated in Article 65 and the criminal

provisions of Articles 67 to 69 of the PDP Law. Fifth, if non-consensual deepfakes are directed at women in the form of degrading sexual content, the provisions of the TPKS Law regarding electronic-based sexual violence can serve as additional reference, particularly regarding the criminalization of the distribution of sexual content without consent, known as online gender-based violence (KBGO).

Norm Gaps and Overlaps

The mapping above shows that Indonesian positive criminal law actually has several instruments that can indirectly cover non-consensual deepfake acts. However, at least three fundamental gaps were identified. First, there is no single provision that explicitly formulates the element of non-consent as a separate element of the crime, even though this element is the core of the moral and legal issues surrounding deepfakes, as emphasized in the study regarding the lack of specific regulations regarding the misuse of artificial intelligence in the form of pornographic deepfakes. Second, there is overlapping norms between the ITE Law, the Pornography Law, the New Criminal Code, the PDP Law, and the TPKS Law, which has the potential to raise questions regarding the principle of *lex specialis derogat legi generali* and the risk of double jeopardy or vice versa, uncertainty regarding which norm should be applied first. Third, the existing provisions are generally oriented towards the dissemination stage of content, while the production or creation stage of the synthetic image itself, regardless of whether it is then disseminated or not, has not been expressly qualified as a stand-alone criminal act, even though the process of making it alone has given rise to a violation of the victim's self-image autonomy.

This gap aligns with the conclusions of several legal studies by practitioners, which state that, despite several regulatory references based on the type of crime, Indonesia does not yet have specific regulations that comprehensively address the misuse of artificial intelligence technology in the form of deepfakes. Therefore, existing law enforcement remains reactive and case-based. This situation implies a lack of legal certainty (*rechtszekerheid*) for both law enforcement officers and victims, as the process of proving the elements of the crime must be constructed analogously from norms originally designed for conventional crime contexts.

Non-Consensual Deepfakes as a Human Rights Violation

From a human rights perspective, non-consensual deepfakes touch on at least three dimensions of fundamental rights. First, the right to privacy, which is guaranteed by Article 28G paragraph (1) of the 1945 Constitution, which affirms that everyone has the right to protection of themselves, their families, their honor, their dignity, and their property under their control. At the international level, similar guarantees are found in Article 12 of the Universal Declaration of Human Rights and Article 17 of the ICCPR, which prohibit arbitrary interference with a person's private life. Non-consensual deepfakes, especially those of a sexual nature, constitute a unique form of privacy violation because they involve not only the collection or dissemination of personal data, but also the creation of a new synthetic reality that the victim never intended or experienced, thus attacking an individual's control over their own representation in the public sphere.

Second, the right to honor and dignity, which is also guaranteed in Article 28G paragraph (1) of the 1945 Constitution and Article 17 of the ICCPR. Sexually charged deepfake content that superimposes the victim's face onto another person's body creates a social stigma as if the victim was actually involved in the fabricated activity, even though it never actually occurred. This reputational damage is irreversible because digital content can continue to circulate and be archived again even after it has been removed from its original platform.

Third, in the context of victims who are predominantly women, sexual deepfakes can be categorized as a form of online gender-based violence (OGBV) that intersects with the right to feel safe and free from violence as guaranteed in the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW) which has been ratified by Indonesia through Law Number 7 of 1984. The nature of OGBV emphasizes that non-consensual deepfakes are not just an information technology issue, but also a structural issue regarding power relations and gender vulnerability in the digital space, so that the criminalization approach needs to consider the victim's perspective as a whole, not only the technical perspective of digital forensic evidence alone.

Brief Comparison with Legal Developments in Other Jurisdictions

Several jurisdictions have moved more progressively in responding to non-consensual deepfakes through explicit criminal norms. The United Kingdom, through the Online Safety Act 2023 and amendments to the Sexual Offences Act 2003, has criminalized the sharing or threat of sharing of digitally manipulated intimate content without consent as a separate offense,

regardless of the perpetrator's intention to disseminate it widely. Several states in the United States, such as California and Texas, have also enacted specific offenses for non-consensual pornographic deepfakes and political deepfakes in the lead-up to elections. The European Union, through its AI Act, has established a disclosure obligation for content generated or manipulated by artificial intelligence, although this approach is more oriented toward technological regulation than direct criminalization. This comparison demonstrates a global trend toward formulating norms that specifically target two key elements: the lack of consent and the synthetic-manipulative nature of content—two elements that have not been explicitly formulated in Indonesian positive criminal law.

Urgency of Special Criminalization and Policy Recommendations

Based on the discussion above, the formulation of specific criminalization of non-consensual deepfakes is urgent for at least three reasons. First, philosophical reasons, namely the protection of self-image autonomy as an integral part of human dignity guaranteed by the constitution. Second, sociological reasons, namely the increasing frequency and scale of misuse of generative artificial intelligence technology in society, which demands a responsive legal response. Third, legal reasons, namely the gap in norms, as explained above, which causes law enforcement against non-consensual deepfakes to still rely on analogical constructions that are vulnerable to legal uncertainty.

Policy recommendations that can be formulated include several things. First, the need to formulate a specific offense regarding "non-consensual digital image manipulation" (non-consensual synthetic image manipulation), which makes the lack of consent the primary element of the offense, regardless of the content being sexual, fraudulent, or defamatory. This norm could be formulated as a *lex specialis* in the upcoming revision of the ITE Law or as an implementing regulation of the New Criminal Code. Second, strengthening victim remedy mechanisms, including the right to obtain rapid content removal (rapid takedown mechanism), the right to reputation restoration, and access to legal and psychological assistance. Third, strengthening the institutional capacity of law enforcement officials in digital forensic evidence of artificial intelligence-engineered content, considering that proving the authenticity or falsity of audiovisual content requires specialized technical expertise. Fourth, cross-sectoral harmonization between the Ministry of Communication and Digital, the Indonesian National Police, and women's and children's protection agencies, to ensure that the handling of non-

consensual deepfake cases is carried out in an integrated manner and with a victim-centered perspective.

CONCLUSION

Based on the analysis, two main points can be concluded. First, Indonesia's current positive criminal law covers non-consensual deepfakes partially through analogous construction of several sectoral provisions, namely the ITE Law and its amendments, the Pornography Law, the New Criminal Code (specifically Articles 172, 407, and 492), the PDP Law, and the TPKS Law, without any norms that explicitly formulate the elements of lack of consent and the synthetic nature of a person's identity as separate elements of the offense. This condition creates gaps and potential overlapping norms that impact legal uncertainty in their enforcement. Second, from a human rights perspective, non-consensual deepfakes constitute a violation of the right to privacy, the right to honor and dignity, and have the potential to constitute a form of online gender-based violence. Therefore, their handling cannot be approached solely from a technical-technological perspective, but must be based on the comprehensive protection of victims.

Therefore, a specific criminalization of non-consensual deepfakes is needed in the form of *lex specialis* that explicitly defines the lack of consent as the core of the offense, accompanied by strengthening victim redress mechanisms and institutional harmonization between law enforcement agencies. Further research is recommended to empirically examine the effectiveness of law enforcement in non-consensual deepfake cases handled by law enforcement officials, in order to enrich the database for formulating criminal law policies that are more responsive to developments in artificial intelligence technology.

BIBLIOGRAPHY

- AHL Law Office. (2026). Deepfake Fraud and Criminal Law Traps in Indonesia: A Complete Guide 2026. Retrieved from <https://ahl-lawoffice.com>.
- Chesney, R., & Citron, D. K. (2019). Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *California Law Review*, 107(6), 1753-1820.
- European Union. Regulation (EU) 2024/1689 (Artificial Intelligence Act).

- Hukumonline. (2024). What is Deepfake Porn and the Criminal Penalties for Perpetrators. Hukumonline Clinic. Accessed from <https://www.hukumonline.com>.
- Kopusindo Journal. Legal Analysis of Criminal Liability of Pornographic Deepfake Perpetrators. Journal of Law and Public Policy Studies. Accessed from <https://jurnal.kopusindo.com>.
- United Nations. (1948). Universal Declaration of Human Rights.
- United Nations. (1966). International Covenant on Civil and Political Rights.
- Republic of Indonesia. The 1945 Constitution of the Republic of Indonesia.
- Republic of Indonesia. Law Number 1 of 2023 concerning the Criminal Code.
- Republic of Indonesia. Law Number 11 of 2008 concerning Electronic Information and Transactions as amended by Law Number 19 of 2016 and Law Number 1 of 2024.
- Republic of Indonesia. Law Number 12 of 2022 concerning Criminal Acts of Sexual Violence.
- Republic of Indonesia. Law Number 27 of 2022 concerning Personal Data Protection.
- Republic of Indonesia. Law Number 44 of 2008 concerning Pornography.
- Republic of Indonesia. Law Number 7 of 1984 concerning Ratification of the Convention on the Elimination of All Forms of Discrimination against Women (CEDAW).
- SIP Law Firm. (2025). Criminal Law Enforcement of Deepfake Content. Accessed from <https://siplawfirm.id>.
- United Kingdom. Online Safety Act of 2023.
- United Kingdom. Sexual Offenses Act 2003 (as amended).
- Muhammadiyah University of Makassar. (2026). Immoral Deepfake Case, Komdigi Ready to Cut Off Access X, Unismuh Lecturer Highlights Digital Human Rights Violations. Accessed from <https://hukumbisnis.unismuh.ac.id>.