RECHTSVINDING
Vol. 4, 1 (January-June, 2026), pp. 353-358
EISSN: 2987-4424, DOI: https://doi.org/10.59525/rechtsvinding.1541

Reconstruction of Criminal Liability for Deepfake Misuse in Defamatory Defenses in Indonesia

Hairul Saleh Satrul[1], Nur Hafizal Hasanah[1]

[1] Makassar State University, Makassar, Indonesia

ARTICLE INFO

Keywords:

Deepfake;
Defamation;
Criminology

Article history:

Received 2026-02-22
Revised 2026-03-26
Accepted 2026-04-30

ABSTRACT

The deepfake phenomenon as a product of artificial intelligence (AI) has created a new dimension in the crime of defamation. This research aims to analyze the deepfake phenomenon through criminological and juridical approaches to map the effectiveness of regulations in Indonesia. The method used is normative juridical research with a conceptual and case study approach. The results show that deepfake technology facilitates attacks on honor through highly realistic visual manipulation, where 96% of deepfake content on the internet is non-consensual pornography. Juridically, Article 27A of the ITE Law No. 1 of 2024 and the New Criminal Code provide a basis for enforcement, but there is still a functional legal vacuum regarding the liability of technology developers. The implications of this research emphasize the need for specific regulations (sui generis) and strengthening digital forensics to overcome evidentiary challenges in court.

Corresponding Author:
Hairul Saleh Satrul
Makassar State University, Makassar, Indonesia; hairul.saleh.satrul@unm.ac.id

1. INTRODUCTION

Advances in information technology in the era of digital transformation have massively accelerated the use of Artificial Intelligence (AI) in modern society (Sulubara, Tasril, & Nurkhalisah, 2025). One of the most controversial manifestations of this development is deepfake technology, a deep learning-based method that is capable of engineering audio-visual content with a very high level of realism (Nguyen et al., 2022). This technology is basically an evolution of traditional digital image manipulation, howeverhas now reached a level where fabrication is difficult to distinguish from objective reality.

Technically, deepfakes operate through an architecture of Generative Adversarial Networks (GANs), where two AI algorithms conflict with each other: a generator that creates fake data and a discriminator that detects fakes. This technical process allows facial engineering (face swapping), lip movements (lip-syncing), and voice cloning (voice cloning) which is very persuasive. Ease of access to applications like FakeApp has democratized high-level manipulation capabilities for the general public.

The historical roots of this manipulation can be identified back to the Video Rewrite software in 1997, but the name "deepfake" was only popularized in 2017 via the Reddit platform. Since then, the

use of this technology has rapidly expanded beyond entertainment purposes and has begun to spread into the criminal realm. The emergence of a "crisis of trust" or epistemic crisis is a logical consequence, in which the public losescertainty to believe what they see.

The wild's dividend phenomenon poses a new criminological threat, where criminals can refute original video evidence by claiming it to be a deepfake (Cavalli, 2024). This uncertainty undermines the overall integrity of visual information in the digital space. This sparked public unrest as an individual's reputation can now be destroyed in a matter of seconds through synthetic content.

Statistics from Sensity AI show a grim reality, where around 96% of deepfake content on the internet is non-consensual pornography (Cavalli, 2024). The majority of victims in this category are women, whose social dignity is exploited through unauthorized visual manipulation. This attack is not just a violation of privacy, but a form of online gender-based violence (KBGO) that destroys the victim's bodily autonomy.

In the political field, deepfakes have been used as an instrument of disinformation to destabilize global democracy. Michal's fake audio case Šimečka in the 2023 Slovak Election is a stark warning about how synthetic media can change the public narrative in a short time (Meaker, 2023). In Indonesia, a similar method also emerged by manipulating the faces of public figures to spread fake government aid and hoaxes.

Legal challenges arise because existing regulations are considered not yet fully responsive to the uniqueness of digital evidence from synthetic media. Even though the 2024 ITE Law has been revised, experts have identified a "functional legal vacuum" regarding generative AI technology. Law enforcement often has to make creative interpretations of general articles to deal with this highly technical modus operandi.

The aspect of forensic evidence at trial is also a crucial obstacle in the search for material truth. Deepfake content often has "clean" metadata, making it difficult for traditional verification methods to detect manipulation. The capacity gap between criminals and digital forensic tools create major challenges for the criminal justice system.

Therefore, reconstruction of national legal policies is needed to provide stronger protection for citizens' digital reputations (Cavalli, 2024). Law enforcement must not just stop at general legal texts, but must be able to reach the specifics of AI's modus operandi (Arslan, 2023). Synchronization between the ITE Law, the New Criminal Code, and the Personal Data Protection Law (UU PDP) is the main key (Basah, Wijaya, & Januardy, 2025).

This research aims to analyze the dynamics of these crimes through a criminological lens and evaluate Indonesia's juridical readiness. Through a review of the latest literature and case analysis, this research is expected to contribute to the development of adaptive cyber law policies. Protecting the right to honor in the era of synthetic reality must be the country's top priority.

2. METHODS

This research uses a type of normative legal research (normative lawresearch) which positions law as an applicable rule or norm. The focus of the study lies in the inventory of Indonesian positive law, synchronization of statutory regulations, as well as analysis of legal doctrine related to artificial intelligence crimes. The secondary data used includes statutory regulations (2024 ITE Law, New Criminal Code, PDP Law) as well as leading international scientific literature.

The approach methods applied include the statutory approach, conceptual approach and case approach. A statutory approach is used to examine the relevance of Article 27A of the ITE Law and Article 433 of the New Criminal Code to the deepfake mode. The conceptual approach helps understand criminological theory in explaining deviations in the digital space, while the case approach is used to dissect the impact of real incidents in Indonesia.

Data analysis was carried out qualitatively using the normative-descriptive analysis methodwhich is based on syllogism logic. Legal norms are positioned as major premises which are correlated with legal facts related to synthetic media as minor premises to obtain logical legal conclusions. All research

results are presented systematically in order to answer problems regarding the criminal responsibility of perpetrators and the effectiveness of AI crime victim protection mechanisms.

3. FINDINGS AND DISCUSSION

3.1 Criminological Characteristics and Crime Patterns of Deepfakes

The criminological phenomenon of deepfake crimes can be explained through Routine Activity Theory, which states that crimes occur when motivated perpetrators meet, suitable targets, and the absence of capable protectors. In cyberspace, "capable defenders" often fail due to the limitations of platform detection algorithms in identifying synthetic media in real-time (Prayoga & Tuasikal, 2025). This is exacerbated by the social learning environment (social learning theory) in anonymous online forums that makes it easy totransmission of technical knowledge of manipulative content creation (Wibowo, Wangsajaya, & Surahmat, 2023).

The motivation of perpetrators of these crimes is often multidimensional, ranging from sexual gratification to economic gain through blackmail schemes (Wolfe & Hermanson, 2004). Based on Fraud Diamond Theory, perpetrators have the technical capacity obtained from the digital learning process to manipulate identities for personal gain. A clear example is the spike in cases of voice cloning fraud in Indonesia which will increase rapidly between 2022 and 2023, where perpetrators imitate the voices of executives for illegal financial instructions (Kristiyenda, Faradila, & Basanova, 2025).

Victimology studies reveal that deepfake victims experience suffering that goes beyond material loss, including severe psychological trauma and permanent social stigma (Pasaribu, Saputra, Prayogo, & Taun, 2025). People tend to believe visual evidence more than verbal clarification, resulting in damage to the victim's reputation that is destructive and difficult to recover from. This long-term impact often triggers stress disorderspost-traumatic (PTSD) because victims feel they have lost control of their own visual identity.

In Indonesia, the case of a manipulative video similar to Nagita Slavina (Report LP/B/100.1/2002/SPKT/RESORT JAKPUS/PMJ) is clear evidence of criminological challenges in law enforcement. Even though the video was technically proven to be a manipulation, tracking the initial spreader was hampered by the nature of digital anonymity (Wirogioto & Belgradoputra, 2026). This case shows that the victim's fame actually increases the attractiveness for the perpetrator to carry out character assassination for the sake of content monetization.

The shift in modus operandi from conventional defamation to synthetic media indicates an increase in the technical capacity of digital criminals. Perpetrators are now taking advantage of regulatory gaps and weaknesses in people's digital literacy to carry out very systematic honor attacks. Criminological recommendations emphasize the importance of strengthening critical digital literacy and developing AI detection tools by Cyber Police to suppress future criminogenic city figures.

3.2 Juridical Analysis: Legal Instrument Readiness and Forensic Challenges

Juridically, acts of defamation through deepfakes are accommodated through Article 27A of Law no. 1 of 2024 concerning ITE (Cavalli, 2024). This article prohibits anyone from attacking another person's honor or good name by accusing them of something via an electronic system. The use of the victim's visual identity in a reputation-damaging scenario automatically fulfills the elements of an attack on honor with the intent of making it public knowledge.

Integration with the New Criminal Code (UU No. 1 of 2023) through Article 433 provides an additional layer for law enforcement against insult offenses in general. Even though the ITE Law is lex specialis, the New Criminal Code provides normative standards regarding more modern justification and excuse reasons.

Table 1. Comparison of Foundations Digital Defamation Law

Legal Basis	Related Articles	Main Criminal Elements	Maximum Sanctions
UU ITE No. 1/2024	Article 27A	Attacking honor via electronic media	2 years in prison
Criminal Code No. 1/2023	Article 433	Attacking honour in public	9 months in prison
Criminal Code No. 1/2023	Article 492	Fraud with false identities	4 years in prison
PDP Law No. 27/2022	Article 65	Illegal use of biometric personal data	5 years in prison
Pornography Law No. 44/2008	Article 4	Content creation violates decency	12 years in prison

Evidence challenges remain a major obstacle, with deepfake digital evidence often having traces of metadata that appear original. Conventional forensic verification standards often fail to detect subtle manipulations generated by GANs engines. Therefore, there is an urgency to formulate standardized AI forensic protocols to guarantee legal certainty and material truth in cyber trials.

The victim protection mechanism is also strengthened through the PDP Law no. 27 of 2022, especially regarding misuse of biometric data without the data subject's permission. Victims have the right to request restoration of their good name and deletion of content (right to be forgotten) if they are proven to have been harmed. However, implementing the quick takedown procedure still requires closer technical coordination between the government and global digital platform providers.

A substantive legal vacuum regarding the operational definition of deepfakes in law is often triggeringmultiple interpretations at the investigative level. Without specific legal terminology, law enforcers are forced to use analogies that are vulnerable to exceptions in court. Future regulatory reformulation must adopt the principle of radical transparency to prevent systemic misuse of AI technology (Dearden & Parti, 2021).

4. CONCLUSION

Deepfake crimes represent a fundamental threat to human dignity and reputation in the contemporary digital ecosystem. As an evolved form of conventional defamation, this technology allows attacks on honor to be carried out with a level of visual persuasiveness that surpasses human reasoning.

From a criminological perspective, this phenomenon shows that the democratization of AI technology, without being balanced by social surveillance, will increase the number of perpetrators who take advantage of anonymity. The dominant motivation of the perpetrators, ranging from extortion to political character assassination, confirms that this crime has become a new commodity.

Juridically, thoughThe 2024 ITE Law and the New Criminal Code have provided enforcement articles, the effectiveness of their enforcement is hampered by technical limitations. The absence of an explicit definition of synthetic media creates a functional legal vacuum that makes it difficult to qualify the perpetrator's actions. The challenge of digital forensics is a crucial obstacle in the search for material truth in trials. Deepfake electronic evidence is often able to bypass traditional verification standards, requiring more sophisticated AI forensic protocols.

International cooperation, particularly through ratification of the Budapest Convention, is an urgent need given the transboundary nature of these crimes. Without an international legal umbrella, perpetrators can easily avoid national jurisdiction. The government needs to immediately develop official interpretation guidelines for the ITE Law regarding AI manipulated content. This is important to provide guidance for investigators to prevent misuse of rubber articles. 14 The state must

guaranteecompensation and restitution mechanisms for victims through a restorative justice approach. Digital restoration of one's reputation should be a right that victims can easily access.

Strengthening people's digital literacy remains the most effective first line of defense in mitigating the negative impacts of deepfakes. The public needs to be educated to have a skeptical mindset towards sensational visual content.

Investment in biometric detection technology and automatic watermark embedding by technology developers should be required through regulation. This is to ensure accountability at every stage of synthetic content production.

With a holistic and integrative approach between technology, regulation and education, Indonesia can face the threat of this synthetic reality. The integrity of human dignity must be maintained amidst the complexity of the era of artificial intelligence.

REFERENCES

Arslan, F. (2023). Deepfake Technology : A Criminological Literature Review. *The Sakarya Journal of Law*, *11*(1), 701–720. https://doi.org/10.56701/shd.1293642

Basah, D. A. Y., Wijaya, A., & Januardy, I. (2025). Kriminalisasi Pelanggaran Protokol Digital : Tinjauan Hukum Pidana Terhadap Penyebaran Deepfake di Media Sosial. *INNOVATIVE: Jurnal of Social Science Research*, *5*(4), 386–398. https://doi.org/https://doi.org/10.31004/innovative.v5i4.20258

Cavalli, F. (2024). *The State Deepfake 2024*. Retrieved from https://5865987.fs1.hubspotusercontent-na1.net/hubfs/5865987/SODF 2024.pdf

Dearden, T. E., & Parti, K. (2021). Cybercrime, Differential Association, and Self-Control: Knowledge Transmission Through Online Social Learning. *American Journal of Criminal Justice*, *46*, 935–955. https://doi.org/10.1007/s12103-021-09655-4

Kristiyenda, Y. S., Faradila, J., & Basanova, C. (2025). Pencegahan Kejahatan Deepfake : Studi Kasus terhadap Modus Penipuan Deepfake Prabowo Subianto dalam Tawaran Bantuan Uang. *ALADALAH: Jurnal Politik, Sosial, Hukum Dan Humaniora*, *3*(2), 149–164. https://doi.org/https://doi.org/10.59246/aladalah.v2i4

Meaker, M. (2023). *Slovakia's Election Deepfakes Show AI Is a Danger to Democracy*. wired.com. Retrieved from https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/

Nguyen, T. T., Nguyen, Q. V. H., Nguyen, D. T., Nguyen, D. T., Huynh-The, T., Nahavandi, S., … Nguyen, C. M. (2022). Deep Learning for Deepfakes Creation and Detection: A Survey. *Computer Vision and Image Understanding*, *223*. https://doi.org/https://doi.org/10.1016/j.cviu.2022.103525

Pasaribu, A. S., Saputra, M. R. K., Prayogo, I. R., & Taun. (2025). Analisis Yuridis Perbedaan Kritik Dengan Pencemaran Nama Baik Dalam Kuhp Dan Undang-Undang Nomor 1 Tahun 2024 Tentang Informasi Dan Transaksi Elektronik (ITE). *JURRISH: Jurnal Riset Rumpun Ilmu Sosial, Politik, Humaniora*, *4*(2), 220–232. https://doi.org/https://doi.org/10.55606/jurrish.v4i2.4748

Prayoga, H., & Tuasikal, H. (2025). Penyebaran Konten Deepfake Sebagai Tindak Pidana: Analisis Kritis Terhadap Penegakan Hukum dan Perlindungan Publik di Indonesia. *Abdurrauf Law and Sharia*, *2*(1), 22–38. https://doi.org/https://doi.org/10.70742/arlash.v2i1.194

Sulubara, S. M., Tasril, V., & Nurkhalisah. (2025). *Perlindungan Hukum Tindak Pidana Cybercrime Dalam Cyberlaw di Indonesia: Perkembangan Tekhnologi dan Tantangan Hukum dalam Mewujudkan Cybersecurity*. Medan: Tahta Media.

Wibowo, A., Wangsajaya, Y., & Surahmat, A. (2023). *Pemolisian Digital dengan Artificial Intelligence*. Depok: Rajawali Pers.

Wirogioto, A. J., & Belgradoputra, R. J. S. (2026). *Memahami Hukum Siber*. CV Intelektual Writer.

Wolfe, D., & Hermanson, D. (2004). The Fraud Diamond: Considering the Four Elements of Fraud. *The CPA Journal*, *74*, 38–42. https://doi.org/https://digitalcommons.kennesaw.edu/cgi/viewcontent.cgi?article=2546&context=facpubs