International Journal of Law
www.lawjournals.org
ISSN: 2455-2194
Received: 16-01-2026, Accepted: 15-02-2026, Published: 02-03-2026
Volume 12, Issue 1, 2026, Page No. 374-377

Criminal sanctions for perpetrators of deepfake technology abuse in defamation on social media

Made Indah Naraswari, I Nyoman Gede Sugiartha, Ni Made Sukaryati Karma
Faculty of Law, Warmadewa University, Denpasar, Indonesia

Abstract
Criminal sanctions for perpetrators of misuse of deepfake technology in criminal acts of defamation on social media. The rapid development of Artificial Intelligence (AI) technology, especially deepfake, is often misused to enable the creation of fake visual or audio content that is made very realistic, so that it has the potential to damage an individual's reputation massively and instantly. With this deepfake, perpetrators can easily manipulate content on social media. The main problems in this study are 1) how to regulate the misuse of deepfake technology and 2) criminal sanctions for the misuse of deepfake technology in Indonesian positive law. The research method used is normative legal research with a statutory and conceptual approach. The results of the study show that the misuse of deepfakes for defamation is classified as a criminal act that attacks a person's honor or good name through electronic information. Criminal sanctions for perpetrators refer to Article 27A of the ITE Law regarding attacks on honor and good name, which carries a threat of two years' imprisonment or a fine of four hundred million rupiah, as well as Article 433 paragraph (2) of the Criminal Code regarding written or pictorial defamation, with a maximum prison sentence of 1 year and 6 months. This study recommends the need to strengthen specific legal sanctions to address the increasingly sophisticated challenges of digital content identification.

Keywords: Sanctions, criminal, social media

Introduction

The fundamental foundation of state administration in Indonesia is the principle of the rule of law (rechtsstaat), expressly stated in Article 1 paragraph (3) of the 1945 Constitution of the Republic of Indonesia. This principle mandates that the state is obliged to guarantee legal certainty, justice, and protection of the rights of every citizen. In today's digital era, this obligation faces increasingly complex challenges with the emergence of technology and social media.

The development of information and communication technology has brought humanity into the era of the digital revolution. Social media such as Instagram, Twitter, Facebook, and TikTok have become public spaces that enable the exchange of ideas and self-expression. This current technological development has given rise to Artificial Intelligence (AI) [1].

Artificial Intelligence (AI) has become one of the major innovations in human civilization. The concept of AI was first introduced by American mathematician John McCarthy in 1955. Today, AI has developed rapidly and penetrated various sectors of life. Many major social media platforms have used AI technology to develop facial recognition and other features [2].

As AI advances, it has given rise to deepfake technology. Deepfake technology is used to manipulate or manipulate a person's audio and video to make them appear highly realistic [3]. The most concerning misuse of deepfake technology is for defamation. Reputations and honor can be damaged through the distribution of deepfake content [4].

The impact of defamation through deepfakes, which possess a high degree of realism, can deceive the public, and their rapid spread can damage a person's reputation [5]. An example of the misuse of deepfake technology is the video of the Minister of Finance, Sri Mulyani. In the video, she appears to state that teachers are a burden on the state. Responding to this, her spokesperson, Deni Surjantoro, stated that the video was not real and was a hoax. This statement was reinforced by Sri Mulyani in a post on her personal Instagram account, confirming that she never said such things.

Sri Mulyani's case represents a clear threat to the honor of public officials and individuals. There are challenges in identifying perpetrators who use anonymous accounts and conceal their digital footprints. Although the perpetrators used AI technology to manipulate audio-visuals to appear realistic, the substance of their actions was to disseminate electronic information accusing them of something for public disclosure. Therefore, such content meets the elements of an attack on honor or reputation, as the fabricated narrative could fuel negative public perceptions of Sri Mulyani's integrity.

The spread of deepfake content has caused public unrest and presented new challenges in criminal law enforcement. In Indonesia, the use of deepfakes has the potential to have negative impacts, particularly in cases of fraud and defamation. The public is increasingly vulnerable to videos using deepfake technology, where individuals or organizations can easily create misleading content [6].

The growth of deepfake content online has increased dramatically year after year, indicating that this technology is becoming increasingly accessible and easily misused by the public. In response to this threat, although Indonesia has legal instruments to address cybercrime and defamation, specific regulations are needed to prosecute perpetrators of deepfake abuse.

From a constructivist perspective, deepfake technology not only produces misleading content but also shapes how society interprets truth through intersubjective constructions of meaning. When the public believes manipulated content to be reality, the line between fact and fiction becomes blurred. The danger of deepfakes lies not only in the element of deception but also in their potential to undermine the social fabric of reality [7].

Law Number 1 of 2024 concerning the Second Amendment to Law Number 11 of 2008 concerning Electronic Information and Transactions, Article 27A, which replaces Article 27 paragraph (3), prohibits attacking a person's honor or reputation by making accusations publicly known in the form of electronic information. Furthermore, the crime of defamation is also regulated under Article 310 of the Criminal Code, which has now been transformed into Article 433 of Law Number 1 of 2023 concerning the Criminal Code.

The challenges that arise in identifying and verifying deepfake content are that the general public often has difficulty distinguishing between genuine and doctored videos. Limited resources and technical expertise also hinder investigations into cases involving deepfakes. As a result, perpetrators can potentially escape prosecution, while victims suffer significant material or immaterial losses.

The phenomenon of misuse of deepfakes for defamation creates a mismatch between the pace of technological development and legal adaptation. Based on the above description, this study aims to analyze the regulations governing the misuse of deepfake technology for defamation on social media and to discuss sanctions for perpetrators who misuse deepfake technology for defamation on social media.

Method

The normative data search is based on laws and regulations that emphasize the honor of a person's good name and theoretical investigation related to this composition. This is carried out through a process of discovering legal rules, legal principles, and legal doctrines to answer the legal problems faced [8]. The sources of legal material for this research are primary legal sources derived from laws and regulations, as well as secondary legal sources derived from legal books and journals to complete the theory without deviating from positive law. The technique for collecting legal material is carried out through library studies by reading and reviewing various legal literature. The analysis of legal material is carried out using a descriptive analysis method that describes laws and regulations and legal concepts related to the problem being studied.

Results and Discussion

1. Regulation of the Misuse of Deepfake Technology in Criminal Acts of Defamation on Social Media

Given the diversity of individual perspectives on issues, expressing opinions is an integral part of social interaction. The right to express opinions and thoughts freely is guaranteed to all Indonesians as a form of freedom of expression. In this modern era, this freedom of opinion is easily expressed through social media platforms.

Social media is a digital platform that facilitates broad communication and interaction for every individual. Social media platforms such as Facebook, Twitter, Instagram, TikTok, and YouTube have become primary platforms for expression, sharing information, and building communities. However, their presence has fueled an increase in cases of digital insults or defamation. Defamation is an act of attacking a person's honor or good name, resulting in the destruction of the victim's reputation. The Criminal Code (KUHP) defines seven types of defamation, including defamation (Article 433), slander (Article 434), minor insults (Article 436), slanderous complaints (Article 437), false accusations (Article 438), and defamation of the deceased (Article 439). Among these types of defamation, the misuse of deepfake technology is classified as a criminal act of defamation, whether in writing or through images, as stipulated in Article 433 paragraph (2) of the Criminal Code.

Defamation on social media currently often occurs in the form of videos, derogatory comments, status updates, and other posts on social media. Violations of reputation through image and audio-visual manipulation significantly impact social life due to the speed and unlimited reach of their distribution [9].

Defamation is a dynamic phenomenon, where legal protection for such incidents guarantees the right of every individual to have their dignity respected. Honor refers to an individual's right to be respected in a social environment, while good name relates to a person's moral reputation. Both aspects are measured by the prevailing norms and values of the society in which the incident occurred [10].

Criminal regulations need to adapt to the unique nature of digital platforms to prevent the misuse of the law as a tool of silencing. The Criminal Code (KUHP) serves as the lex generalis and the Electronic Information and Transactions (ITE) Law (Lex specialis) are needed to handle defamation cases. The defamation provisions in the ITE Law, which address anyone who intentionally attacks another person's honor or reputation by making an accusation for public disclosure in the form of electronic information, are stipulated in Article 27A.

Defamation under Article 27A is categorized as an absolute complaint offense. This offense requires the victim to personally report the loss they have suffered. The reporting period is 6 months if in Indonesia and 9 months if outside Indonesia, which is rooted in Article 24 of the Criminal Code. This emphasizes the authority of the police to act if there is a complaint from a party who feels aggrieved [11].

The relationship between the Criminal Code and the Electronic Information and Transactions Law (ITE Law) governing defamation is a manifestation of the legal principle of lex specialis derogate legi generali. The ITE Law is an extension of the material offenses contained in the Criminal Code. To understand and interpret the element of "attacking honor or good name," the ITE Law serves as a special criminal procedure law for material offenses contained in the Criminal Code [12]. This codification demonstrates that the element of attacking honor or good name in cyberspace has a standard similar to that in offline space, where the criminal aspect remains based on the offense of complaint.

2. Sanctions Against Perpetrators of Deepfake Technology Abuse in Defamation on Social Media

Sanctions are a coercive tool to ensure legal rules are complied with, not merely beautiful formulas on paper. Without firm sanctions, legal norms lose their coercive power and are unable to fulfill their function as social control [13]. The application of sanctions serves to ensure the achievement of state goals; sanctions act as a coercive instrument to ensure public compliance with applicable regulations.

Sanctions to punish an individual are only used when absolutely necessary to maintain public order. Punishment is considered a crime that is forced to be committed only to control unstable human behavior [14]. Justice and truth in the

law require an authoritative oversight system to maintain its upholding [15]. Based on Article 64 of the Criminal Code, the classification of criminal sanctions includes Principal Criminal Offenses, Additional Criminal Offenses, and Special Criminal Offenses.

Technically, deepfakes process existing data to create new content, such as superimposing a person's face into a video where they never actually participated. The misuse of deepfake technology has spread to various sectors, from politics and economics to pornographic content [16]. To prosecute perpetrators, focused regulations, namely the ITE Law and the Criminal Code, can be used.

The Indonesian legal system currently faces a lack of specific regulations governing deepfakes. Amid the lack of specific regulations, law enforcement against the misuse of this technology must continue to refer to existing basic principles, one of which is the principle of geen straf zonder Schuld (no crime without fault) as the basis for criminal liability [16].

Article 433 of the Criminal Code in conjunction with Article 27A of the ITE Law is used to address defamation committed by perpetrators of deepfake technology abuse. Article 45 paragraph (4) of the ITE Law discusses sanctions for violating Article 27A, which carries a two-year prison sentence or a fine of four hundred million rupiah. Deepfakes are associated with data manipulation, as defined in Article 35 in conjunction with Article 27A. Article 51 paragraph (1) of the ITE Law imposes a criminal penalty of twelve years' imprisonment or a fine of twelve billion rupiah. This article is more in line with the crime of defamation in the Criminal Code. In the case of deepfakes, the perpetrator who creates fake content that damages someone's reputation and distributes it on social media clearly fulfills the elements of attacking honor, accusing someone of something, and so that it becomes public knowledge. This article is an absolute complaint offense, meaning prosecution can only be carried out upon a victim's complaint.

If the deepfake content created and distributed contains content that violates morality, the perpetrator can be charged under Article 27 paragraph (1) in conjunction with Article 45 paragraph (1) of the Electronic Information and Transactions (ITE) Law, which carries a penalty of six years' imprisonment or a maximum fine of one billion rupiah.

In addition to the ITE Law, which imposes sanctions, the Criminal Code also contains sanctions related to the crime of defamation. The current Criminal Code uses a fine category system stipulated in Article 79 paragraph (1). Article 433 paragraph (2) of the Criminal Code, concerning defamation through writing or images, carries a penalty of one year and six months' imprisonment or a category III fine of fifty million rupiah.

Perpetrators who misuse deepfake technology, in addition to being subject to sanctions under these articles, also include an article in the Criminal Code that regulates the qualifications for deepfakes, namely Article 492 concerning fraud. In the context of technological advancements, this article is relevant for prosecuting the misuse of deepfake technology, as it is categorized as a digital deception tool capable of creating convincing false visual or audio representations. Article 492 of the Criminal Code provides for a criminal penalty of four years' imprisonment or a Category V fine of five hundred million rupiah, in accordance with Article 79 paragraph (1) of the Criminal Code.

Perpetrators often use anonymous accounts, virtual private networks, or operate from other jurisdictions, making it extremely difficult to trace and identify the true identities of the creators and first distributors. The perpetrators, victims, and social media platforms can be located in different countries. As technology advances, deepfakes are becoming increasingly difficult to distinguish from genuine videos [17].

Conclusion

Based on the research results, it can be concluded that defamation is defined as an act of attacking someone's honor or good name by accusing them of something so that it becomes public knowledge. Regulations regarding the crime of defamation are included in Article 433 of the Criminal Code and Article 27A of Law Number 1 of 2024 concerning Electronic Information and Transactions. Deepfake is categorized as an insult in the form of electronic information that attacks someone's honor. The sanctions that can be given to the perpetrator are based on Article 45 paragraph (4) of the ITE Law in the form of two years' imprisonment or a fine of four hundred million rupiah, and if it is associated with data manipulation, then twelve years' imprisonment or a fine of twelve billion rupiah is imposed based on Article 51 paragraph (1) of the ITE Law. If the deepfake content contains immoral content, then Article 27 paragraph (1) in conjunction with Article 45 paragraph (1) of the ITE Law provides a criminal sanction of six years' imprisonment or a fine of one billion rupiah. Meanwhile, in the Criminal Code, the sanction for defamation through images or writing is one year and six months' imprisonment or a category III fine. And regarding deepfake content that aims to deceive by manipulating identity, Article 492 of the Criminal Code provides a criminal sanction of four years' imprisonment or a category V fine of five hundred million rupiah.

References

1. Darmawan, Aang Kisnu. Social Media Analytics: Konsep Dan Penerapannya Dengan Rapid Miner/Orange. Banten. Yayasan Pendidikan dan Sosial Indonesia Maju, 2022.
2. Sijabat, Sarah Amanda Uly, Diana Lukitasari. “Konten Gambar Dan Video Pornografi Deepfake Sebagai Suatu Bentuk Tindak Pidana Pencemaran Nama Baik.” RECEIDEVE: Jurnal Hukum Pidana Dan Penanggulangan Kejahatan,2024:13(2):179–94. https://doi.org/10.20961/recidive.v13i2.86771
3. Santoso, Joseph Teguh. Kecerdasan Buatan (Artificial Intelligence). edited by M. Sholikan. Yayasan Prima Agus Teknik, 2024.
4. Alam, Wawan Tunggul. Pencemaran Nama Baik Di Kehidupan Nyata & Dunia Internet. Wartapena, 2016.
5. Suhariyanto Budi. Sistem Pemidanaan Terhadap Pelaku Tindak Pidana Penghinaan Dan Pencemaran Nama Baik Melalui Sarana Teknologi Informasi. Pertama. Jakarta: Kencana, 2019.
6. Sih Yuliana Wahyuningtyas. Pelindungan Data Biometrik Dalam Pemrosesan Oleh Artificial Intelligence (AI) Untuk Teknologi Deepfake. Jakarta: Atma Jaya, 2024.
7. Nurdin, Sri Wahyuni, Imam Fadhil Nugraha. “Ancaman Deepfake Dan Disinformasi Berbasis Ai: Implikasi Terhadap Keamanan Siber Dan Stabilitas Nasional Indonesia.” JIMR: Journal Of International

Multidisciplinary Research,2025:4(1):73. doi: https://doi.org/10.62668/jimr.v4i01.1551.
8. Efendi, Jonaedi. Metode Penelitian Hukum Normatif Dan Empiris. 5th ed. Jakarta: Kecana, 2022.
9. Puteri, Irene, Alfani Sofia, Catrina Yuka, Shabrina Aurellia, Nafisah Desuardi, and Mirelle Elicia Perera. "Analisis Hukum Terhadap Pencemaran Nama Baik Di Media Sosial Dalam Perspektif Perbuatan Melawan Hukum." Jurnal Ilmu Hukum, Humaniora, Dan Politik,2025:6(2):06–13. https://doi.org/10.38035/jihhp.v6i2.
10. Putu Angga, Made Arjaya dan Sukaryati Karma. "Peranan Alat Bukti Elektronik Dalam Tindak Pidana Pencemaran Nama Baik." Jurnal Interpretasi Hukum,2021:2(3):602–6. https://doi.org/10.22225/juinhum.2.2.3452.422-428.
11. Pasca, Rezkyta, Abrini Daeng, Sigid Suseno, Budi Arta Atmaja. "Penerapan Pasal 27 Ayat (3) Undang-Undang ITE Dalam Perkara Pencemaran Nama Baik Melalui Media Sosial Terhadap Kelompok Orang Application of Article 27 Paragraph (3) of the ITE Law in Cases of Defamation through Social Media against Groups of People." Jurnal Fundamental Justice,2022:3(1):19–35. doi: https://doi.org/10.30812/fundamental.v2i2.1796.
12. Andini, Orin Gusta. "Revaluasi Tindak Pidana Pencemaran Nama Baik." Jurnal Jatiswara,2019:34(2):143–54.
13. Rahardjo, Satjipto. Ilmu Hukum. Delapan. Bandung: PT Citra Aditya Bakti, 2016.
14. Pande Juli Artana, I Nyoman Gede Sugiartha. I. Made Minggu Widyantara. "Sanksi Pidana Terhadap Pelaku Pencemaran Nama baik Melalui Media Sosial (Studi Kasus Putusan Pengadilan Nomor). Jurnal Interpretasi Hukum,2022:3(1):25–30. https://doi.org/10.22225/juinhum.3.1.4633.25-30.
15. Bakhri, Syaiful. Hukum Sanksi di Berbagai Praktek Peradilan. Jakarta: UM Jakarta Press, 2020.
16. Hapid, Fasa Muhamad, Ija Suntana, and Muhammad Yayan Royani. "Penerapan Asas Geen Straf Zonder Schuld Dalam Penindakan Terhadap Kejahatan Penyalahgunaan Teknologi Deepfake Application of the Geen Straf Zonder Schuld Principle in Taking Action Against Crimes of Misuse of Deepfake Technology." Jurnal USM Law Review,2024:7(3):4–7. doi: https://doi.org/10.26623/julr.v7i3.9686.
17. Chazawi, Adami. Tindak Pidana Informasi & Transaksi Elektronik: Penyerangan Terhadap Kepentingan Hukum Pemanfaatan Teknologi Informasi Dan Transaksi Elektronik. Pertama. Malang: Media Nusa Creative, 2018.