JURNAL AKTA
Volume 12 No. 2, June 2025
SINTA 2,
Decree No. 164/E/KPT/2021

# Criminal Policy Toward the Crime of Defamation in Cyberspace Through the Use of Deepfake AI

**Wenggedes Frensh**
Faculty of Law, Universitas Medan Area, Medan, Indonesia, E-mail: wenggedesfrensh@staff.uma.ac.id

***Abstract.*** *Currently, humanity is experiencing rapid advancements in science and technology. Information and communication technology has undergone significant developments, particularly with the emergence of Artificial Intelligence (AI). This progress has ushered in the digital era, where people interact in cyberspace through social media platforms that utilize AI. While the use of AI in social media brings positive impacts by enhancing interactions in cyberspace, it also presents the potential for cybercrimes, including defamation through AI-based technology known as Deepfake. Several notable cases of defamation using Deepfake AI have occurred, such as the case of Raffaela Spone in the United States, who sent defamatory Deepfake content targeting members of the Victory Vipers cheerleading team; the manipulated video of Ukrainian President Volodymyr Zelenskyy; the Deepfake video of Indian actress Rashmika Mandanna wearing revealing clothing; and cases involving manipulated Deepfake videos of Indonesian public figures such as Prabowo Subianto, Gibran Rakabuming Raka, and Sri Mulyani. In Indonesia, there is still a lack of clear regulation concerning defamation through the use of Deepfake AI and how to address such crimes through criminal policy. This study aims to examine efforts to combat cyber defamation using Deepfake AI by applying both penal and non-penal criminal policy approaches. The research employs a normative juridical method, using secondary data sources and deductive analysis. The findings show that efforts to address the crime of cyber defamation using Deepfake AI can be implemented through penal policy using formulation strategies, and through non-penal policy by eliminating criminogenic factors contributing to cyber defamation with Deepfake AI. Penal policy countermeasures can be pursued under Law No. 11 of 2008 in conjunction with Law No. 19 of 2016, specifically Article 27 paragraph (3). Non-penal approaches involve the use of Deepfake AI detection technologies, protection of digital identities, digital literacy education, and the enhancement of personal data security.*

***Keywords:*** *Criminal; Cyberspace; Deepfake; Defamation; AI.*

## 1. Introduction

Humans and science continue to evolve in their lives. In their interactions, humans move alongside science to meet their needs. The development of science gives rise to technology that facilitates human daily life in fulfilling those needs. One of the technologies that has seen very significant and rapid growth is information and communication technology. Information and communication technologies such as computers, laptops, and mobile phones are used by society to connect to the internet, creating a new space known as cyberspace.(Kasita, 2022). In Indonesia, internet users are numerous and continue to increase significantly. According to a survey by APJII (Indonesian Internet Service Providers Association), in 2024, internet users in Indonesia grew to over 221 million people, representing about 79.5% of the population. This shows a significant increase from 88.1 million users in 2014. The increase in internet use in Indonesia is due to society's growing need for information and communication in cyberspace. Interaction in cyberspace occurs because society now lives in the era of information and communication technology, where most activities undergo digitalization. According to a survey by We Are Social, around 49.9% or 139 million of Indonesia's population were active social media users in 2024. The most popular platforms include WhatsApp, Instagram, Facebook, TikTok, Telegram, X (Twitter), Facebook Messenger, Pinterest, Snack Video, and LinkedIn. This increase in internet and social media usage clearly illustrates that cyberspace has become an essential space for social interaction and daily activities in Indonesia’s digital era.(Rizki Kurniarullah et al., 2024).

Interactions in cyberspace are conducted using software applications on laptops or smartphones. Some applications have developed rapidly, such as those used to modify images, videos, and audio. These image, video, and audio editing apps are popular because they are seen as entertaining and useful in professional work. These apps, often referred to as Deepfake AI tools, include MaxStudio Ghibli Filter AI Face Swap, Make Your Own AI Influencer, Deepfakes Web, DeepSwap.ai, D-ID, Free Deepfake Maker, and many more. While these tools can serve entertainment and work purposes, they can also be exploited for legal violations and cybercrimes (Pokhrel, 2024).
Interactions in cyberspace are conducted using software applications on laptops or smartphones. Some applications have developed rapidly, such as those used to modify images, videos, and audio. These image, video, and audio editing apps are popular because they are seen as entertaining and useful in professional work. These apps, often referred to as Deepfake AI tools, include MaxStudio Ghibli Filter AI Face Swap, Make Your Own AI Influencer, Deepfakes Web, DeepSwap.ai, D-ID, Free Deepfake Maker, and many more. While these tools can serve entertainment and work purposes, they can also be exploited for legal violations and cybercrimes.

Cybercrime refers to crimes that occur on the internet using information and communication technology as the medium. In Indonesia, during 2024, there were 3,331 recorded cybercrime cases, including fraud, defamation, pornography, hoaxes, hate speech, hacking, and data theft. Many of these cases involve defamation, which occurs because cyberspace facilitates communication and interaction. Defamation in cyberspace is increasingly sophisticated with the help of evolving technologies. Deepfake AI, which uses artificial intelligence to manipulate videos, images, and audio to look authentic, makes defamation cases harder to resolve. Deepfake defamation blurs the line between real and manipulated electronic information and documents, as AI-generated content often appears extremely realistic.(Hukom et al., 2025). Several countries have reported cases of defamation using Deepfake AI. In the United States in 2021, Raffaela Spone from Pennsylvania created and distributed Deepfake images and videos of teenage girls on the Victory Vipers cheerleading team, depicting them nude, drinking, and smoking. In 2022, Ukrainian President Volodymyr Zelenskyy was also a victim when a Deepfake video appeared on social media showing him telling troops to lay down their weapons and surrender to Russia. In 2023, Indian actress Rashmika Mandanna was a victim of Deepfake AI, where a manipulated video showed a woman in revealing clothing that resembled her. In 2025, Indonesia's Cyber Crime Directorate (Dittipidsiber Bareskrim Polri) uncovered a Deepfake scam using the faces and voices of President Prabowo Subianto, Vice President Gibran Rakabuming Raka, and Finance Minister Sri Mulyani.

In Indonesia, defamation using Deepfake AI has become increasingly common among internet users (netizens), raising complex legal challenges. While Indonesia already has legislation related to electronic information and technology  namely, Law No. 11 of 2008, amended by Law No. 19 of 2016 and further amended by Law No. 1 of 2024 it still lacks specific regulations addressing defamation using Deepfake AI. Additionally, there are no significant efforts to reduce criminogenic factors through non-penal policies in tackling Deepfake-related defamation. Therefore, a comprehensive criminal policy is needed to formulate strategies using both penal (criminal law) and non-penal approaches to address these crimes.(Informasi, n.d.). The rise of defamation crimes in Indonesia involving artificial intelligence technology, especially Deepfake, has led to increased manipulation of images and videos on cyberspace platforms and social media. These crimes result in Indonesian citizens becoming victims in cyberspace. Addressing defamation through Deepfake AI requires a specific policy known as criminal policy for cybercrime. There is a need to formulate an integrated criminal policy to combat defamation crimes in Indonesian cyberspace involving Deepfake AI.

## 2. Research Methods

The type of research used is normative juridical research. This research is descriptive-analytical in nature. The data sources are obtained through secondary data consisting of primary legal materials. The data analysis used is qualitative data analysis.

## 3. Results and Discussion

### 3.1. Penal Policy Towards Defamation Crimes Using Deepfake Ai

In efforts to combat crime, G.P. Hoefnagels explains that there are three (3) stages that must be considered: the stage of tackling crime through criminal law (criminal law application), then the stage of crime prevention without imposing punishment (prevention without punishment), and influencing public views on crime and punishment (influencing views of society on crime and punishment/mass media) .Combating cybercrime in the form of defamation using Deepfake AI requires criminal policy as an effort to tackle crime. Criminal law policy refers to crime prevention through legal formulation using legislation. On the other hand, there is also a need for non-penal policies, which are often referred to as preventive measures outside of criminal law. The aim of non-penal policy is to eliminate criminogenic factors that cause defamation using Deepfake AI in cyberspace.(Penanggulangan et al., 2023).

In penal-based policy, there is the formulation policy, which is a legislative policy where law enforcement in abstracto is carried out by legislative bodies. Legislative policy is a plan or program created by lawmakers to address specific problems. The formulation policy as part of legislative policy is a crucial element in the effort to tackle defamation crimes in cyberspace using Deepfake AI. Indonesia already has Law No. 11 of 2008 concerning Electronic Information and Transactions (ITE Law), which was amended by Law No. 19 of 2016, and most recently by Law No. 1 of 2024. Penal criminal policy is reflected in the formulation of the ITE Law as a response to cybercrimes. While defamation in cyberspace is already regulated in the ITE Law, defamation using Deepfake AI presents unique characteristics, as Deepfake AI can produce electronic documents and information that appear authentic when processed using artificial intelligence. Therefore, the current regulation does not specifically address defamation using Deepfake AI. However, defamation involving Deepfake AI as an instrument can still be linked to defamation provisions in the ITE Law.(Frensh, 2022).

Deepfake refers to media content created or manipulated by AI—including videos, audio, or images—designed to mimic reality. The line between what is real and artificial has become so blurred that even experts may struggle to distinguish truth from fabrication without sophisticated tools. Deepfake media is synthesized using artificial intelligence, particularly through a method known as Generative Adversarial Networks (GANs). Types of Deepfake AI manipulation include (a) Face Swapping—superimposing a person's face onto another body in a video, (b) Voice Cloning—replicating someone's voice to produce authentic-sounding speech, and (c) Full Body Synthesis—imitating gestures and body movements to replicate someone's behavior. When Deepfake AI is used to manipulate videos, images, and audio for the purpose of defamation, it constitutes a cybercrime as regulated under the ITE Law.(Yuri S.E Pratiwi & Ravena, 2022).

Defamation using Deepfake AI in cyberspace can be linked to Article 27 paragraph (3) of the ITE Law, which states that any person who intentionally and without authority distributes and/or transmits and/or makes accessible electronic information and/or electronic documents containing defamation and/or insult shall be punished. Article 45 paragraph (3) further stipulates that any person who does so may be sentenced to up to 4 (four) years in prison and/or fined up to IDR 750,000,000 (seven hundred fifty million rupiah). Defamation using Deepfake AI involves manipulating electronic information and/or electronic documents—specifically videos, images, and audio—which clearly fall within the definition of electronic information and documents. Manipulations performed on videos, images, and audio result in electronic information and/or documents that contain insulting content aimed at a person.

Electronic Information refers to one or a set of electronic data, including but not limited to text, sound, images, maps, designs, photographs, electronic data interchange (EDI), emails, telegrams, telex, telecopy, or similar, letters, characters, numbers, access codes, symbols, or perforations that have meaning or can be understood by competent individuals. Electronic Documents refer to any electronic information created, transmitted, sent, received, or stored in analog, digital, electromagnetic, optical, or similar forms, which can be seen, displayed, and/or heard via computers or electronic systems, including but not limited to writings, sounds, images, maps, designs, photographs, or the like, characters, numbers, access codes, symbols, or perforations that have meaning or can be understood by competent individuals. Defamation using Deepfake AI involving videos, images, and audio falls under the category of electronic information and documents as defined in the ITE Law.

According to Article 27 paragraph (3) of the ITE Law, "distributing" means sending and/or spreading electronic information and/or documents to many people or various parties through electronic systems. "Transmitting" refers to sending electronic information and/or documents addressed to a specific party through electronic systems. "Making accessible" refers to any act other than distributing or transmitting that enables electronic information and/or documents to be accessed or known by others or the public. Defamation in cyberspace using Deepfake AI is carried out by sending manipulated videos, images, and audio to social networks or social media. These actions—distributing, transmitting, or making accessible Deepfake AI content with defamatory elements—are criminal acts in cyberspace.

Defamation in cyberspace using Deepfake AI aims to humiliate and attack the honor of an individual through manipulation of videos, images, and audio. Article 27 paragraph (3) of the ITE Law refers to manipulated electronic information and/or documents using Deepfake AI as having defamatory and/or slanderous content. This provision in the ITE Law aligns with the concept of defamation and/or slander as regulated in the Indonesian Penal Code (KUHP). Moeljatno defines defamation as an act of degrading the dignity or reputation of a person in public, whether through words, writing, or actions. Meanwhile, Andi Hamzah defines defamation as delivering false statements that attack the honor or good name of another person, either directly or indirectly Based on the views of Moeljatno and Andi Hamzah, it can be concluded that.(Melani et al., 2020)

### 3.2. Non-Penal Policy Toward the Crime of Defamation Using Deepfake AI

Efforts to combat the crime of defamation involving Deepfake AI must certainly involve technology. The main issue with defamation using Deepfake AI lies in the need to verify whether the videos, images, and audio involved are genuine or manipulated by Deepfake AI. Therefore, it is necessary to have tools that can detect whether such videos, images, and audio are authentic or have been edited by artificial intelligence. These detection tools are crucial for identifying whether a video has been altered using Deepfake AI manipulation [17]. The development of detection tools is important because they incorporate artificial intelligence, just like the tools used to create Deepfake AI content. Many Deepfake AI detection tools learn in the same way as Deepfake AI itself. These tools analyze datasets of real and Deepfake videos and use that data to classify videos as either genuine or Deepfake.(Dolhansky et al., n.d.).

Deepfake AI technology can also be used for identity theft, making it essential to protect users' digital identities in cyberspace. By protecting digital identities—such as securing data, videos, images, audio, and other electronic information or documents—offenders are limited in their ability to commit acts of defamation. Some strategies that can be implemented to protect internet users from becoming victims of Deepfake AI-based defamation include the following steps:(Wang, n.d.)

a) Be cautious when sharing photos, videos, and images;

b) Enable strong privacy settings;

c) Use watermarks on photos;

d) Learn about Deepfake and AI;

e) Use multi-factor authentication;

f) Use long, strong, and unique passwords;

g) Keep your software up to date;

h) Avoid phishing traps;

i) Report defamatory Deepfake content

By protecting digital identities in cyberspace, the misuse of videos, photos, and audio for defamation can be minimized. Protection strategies for digital identity help prevent perpetrators of defamation using Deepfake AI from having the opportunity to commit such acts, as the videos, images, and audio are already secured and cannot be accessed arbitrarily by unknown parties.

Defamation using Deepfake AI often targets internet users who lack knowledge about Deepfake and AI, making it important for users to receive digital literacy education. This education should explain how Deepfake works in entertainment content and how it can be misused for criminal acts such as defamation. As a non-penal policy solution to defamation crimes, a comprehensive digital literacy program is needed. Such a program aims to provide thorough digital literacy education to all groups, ensuring that people of all ages understand that cyberspace is a place where freedom of communication and access to information exists—but within the boundaries of prevailing norms and values. As a result, internet users will avoid creating or sharing Deepfake AI content that could lead to defamation, as they will understand what constitutes legal and illegal behavior. Information and education about how to use AI and Deepfake technology appropriately will help ensure these technologies are used for beneficial purposes(Westerlund, 2019).

The fight against defamation crimes can also be supported by enhancing personal data security. Steps to improve personal data security include:

a) Using strong and unique passwords;

b) Enabling two-factor authentication;

c) Regularly updating software and applications;

d) Using secure Wi-Fi networks;

e) Being aware of phishing attempts;

f) Being cautious about the information shared on social media;

g) Using a password manager;

h) Monitoring accounts regularly

By strengthening personal data security, users of information and communication technology—whether online or offline—are protecting their electronic information and documents. Securing the technology they use, such as laptops, phones, or other devices, ensures that personal data remains protected and cannot be easily exploited by perpetrators of defamation who might use Deepfake AI for manipulation.(Rana & Nobi, 2022)

## 4. Conclusion

The research results indicate that efforts to combat the crime of defamation in cyberspace using Deepfake AI can be carried out through penal policies—specifically formulation policies—as well as non-penal policies that aim to eliminate the criminogenic factors of defamation crimes in cyberspace involving Deepfake AI. Penal policy efforts can be implemented by applying Law No. 11 of 2008 in conjunction with Law No. 19 of 2016, particularly Article 27 paragraph (3). Meanwhile, non-penal policy measures involve the use of Deepfake AI detection technology, protection of digital identity, digital literacy education, and enhancement of personal data security.

## 5. References

### Journals:

Frensh, W. (2022). Kelemahan Pelaksanaan Kebijakan Kriminal Terhadap Cyberbullying Anak Di Indonesia. *Indonesia Criminal Law Review*, *1*(2), 87–99.

Kasita, I. D. (2022). Deepfake Pornografi: Tren Kekerasan Gender Berbasis Online (KGBO) Di Era Pandemi Covid-19. *Jurnal Wanita Dan Keluarga*, *3*(1), 16–26. https://doi.org/10.22146/jwk.5202

Melani, M., Disemadi, H. S., & Jaya, N. S. P. (2020). Kebijakan Hukum Pidana Dibidang Transaksi Elektronik Sebagai Tindak Pidana Non-Konvensional. *Pandecta Research Law Journal*, *15*(1), 111–120. https://doi.org/10.15294/pandecta.v15i1.19469

Pokhrel, S. (2024). Pelindungan Hukum bagi Korban Deepfake Pornografi (Studi

Perbandingan Indonesia dan Korea Selatan). In *Αγαη* (Vol. 15, Issue 1).

Rana, S., & Nobi, M. N. U. R. (2022). Deepfake Detection : A Systematic Literature Review. *IEEE Access*, *10*, 25494–25513. https://doi.org/10.1109/ACCESS.2022.3154404

Rizki Kurniarullah, M., Nabila, T., Khalidy, A., Juniarti Tan, V., Widiyani, H., Hukum Universitas Maritim Raja Ali Haji Abstrak, I., & Kunci, K. (2024). Tinjauan Kriminologi Terhadap Penyalahgunaan Artificial Intelligence: Deepfake Pornografi dan Pencurian Data Pribadi. *Jurnal Ilmiah Wahana Pendidikan*, *10*(10), 534–547. https://doi.org/10.5281/zenodo.11448814

Yuri S.E Pratiwi, V., & Ravena, D. (2022). Kebijakan Restorative Justice terhadap Tindak Pidana Pencemaran Nama Baik di Media Sosial. *Bandung Conference Series: Law Studies*, *2*(1), 1–9. https://doi.org/10.29313/bcsls.v2i1.284

**Books:**

Dolhansky, B., Bitton, J., Pflaum, B., Lu, J., Howes, R., Wang, M., & Ferrer, C. C. (n.d.). *The DeepFake Detection Challenge (DFDC) Dataset*.

Hukom, R., Setiadi, M. H., Ambon, U. D., Ambon, K., Padjadjaran, U., Bandung, K., Barat, J., Info, A., Media, S., Fraud, D., Enforcement, L., & Framework, R. (2025). *Pengaruh Media Sosial terhadap Pola Kejahatan di Era Digital: Studi Kriminologi dengan Pendekatan Netnografi*. *3*(1), 750–768. https://doi.org/10.51903/perkara.v3i1.2353

Informasi, T. (n.d.). *Teknologi informasi*.

Penanggulangan, K., Cyber, K., April, H., & Bu, Y. (2023). *CRIME PENCEMARAN NAMA BAIK DI RUANG SIBER ( Studi Kasus Di Direktorat Reserse Kriminal Khusus Subdit V Cyber Crime Polda Sumut ) SKRIPSI OLEH : HENNY APRIL YANTI BU ' ULOLO UNIVERSITAS MEDAN AREA FAKULTAS HUKUM ( Studi Kasus Di Direktorat Reserse Kriminal Khusus Subdit V Cyber SKRIPSI Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Di Fakultas Hukum Universitas Medan Area*.

Wang, X. (n.d.). *DeepFake Disrupter : The Detector of DeepFake Is My Friend*. 14920–14929.

Westerlund, M. (2019). *The Emergence of Deepfake Technology : A Review*.