

INDONESIA READINESS TO FACE SOCIAL ENGINEERING ATTACKS WITH DEEPFAKE TECHNOLOGY

# a Sayid Muhammad Rifki Noval

**ABSTRACT**

**Theoretical Framework:** In the era of advanced digitalization, concerns about false information and social engineering attacks are currently the main concerns in the field of information security. Information security relies on users' recognition and awareness of the risks associated with social engineering attacks. Identity-protective cognition theory , Information security, social engineering attacks, regulations in Indonesia are used as the underlying theories.

**Purpose:** This research aims to analyze Indonesia’s preparedness to face social engineering attacks with deepfake technology.

**Method:** The research method used in this research is a qualitative method. Activities in data analysis are data reduction, data presentation, and conclusion drawing or verification.

**Result:** Indonesia already has legal regulations that can take action against perpetrators who have abused deepfakes and inflicted losses on their victims, especially if they contain elements of decency.

**Conclusion:** Law enforcement in Indonesia has actually been regulated in the Criminal Code, the Pornography Law, and the Electronic Information and Transaction Law. However, in order to complement its protection and law enforcement, Indonesia can consider the regulations that are being proposed by the United States by specifically regulating deepfakes and also consider media companies to apply content moderation to deepfake material on its platform.

**Keywords:** deepfake, social engineering, revenge porn.

**Received:** 04/09/2023

**Accepted:** 04/12/2023

**DOI:** [https://doi.org/10.55908/sdgs.v11i12.](https://doi.org/10.55908/sdgs.v11i12)727

INDONÉSIA PRONTA PARA ENFRENTAR ATAQUES DE ENGENHARIA SOCIAL COM A TECNOLOGIA DEEPFAKE

# RESUMO

**Estrutura teórica:** Na era da digitalização avançada, as preocupações com informações falsas e ataques de engenharia social são atualmente as principais preocupações no campo da segurança da informação. A segurança das informações depende do reconhecimento dos

a PhD in Law, Islamic Nusantara of University, Indonesia, E-mail: smrn.uninus@gmail.com, Orcid: https://orcid.org/0000-0001-8521-462X



**1**

Miami| v.11, n. 12| pages: 01-17| e0727 |2023. *JOURNAL OF LAW AND SUSTAINABLE DEVELOPMENT*

usuários e da consciência dos riscos associados aos ataques de engenharia social. Teoria da cognição protetora da identidade , segurança da informação, ataques de engenharia social, regulamentos na Indonésia são usados como as teorias subjacentes.

**Propósito:** esta pesquisa visa analisar o preparo da Indonésia para enfrentar ataques de engenharia social com tecnologia de falsificação profunda.

**Método:** O método de pesquisa utilizado nesta pesquisa é um método qualitativo. As atividades de análise de dados são redução de dados, apresentação de dados e conclusão ou verificação.

**Resultado:** a Indonésia já tem regulamentos legais que podem tomar medidas contra os criminosos que abusaram de falsificações e infligiram prejuízos às suas vítimas, especialmente se eles contiverem elementos de decência.

**Conclusão:** A aplicação da lei na Indonésia tem sido regulamentada no Código Criminal, na Lei da Pornografia e na Lei de Transação e Informação Eletrônica. No entanto, a fim de complementar sua proteção e aplicação da lei, a Indonésia pode considerar os regulamentos que estão sendo propostos pelos Estados Unidos, regulando especificamente as falsificações e também considerar as empresas de mídia para aplicar moderação de conteúdo para falsificar material em sua plataforma.

**Palavras-chave:** deepfake, engenharia social, vingança pornô.

# INTRODUCTION

The variety of misuse of technology increasingly shows the complexity of its form and impact. AI, according to John McCarthy, is the technique and science of creating intelligent machines, and there has been significant interest in its potential (Tad et al., 2023). AI has advanced algorithms and impressive decision-making capabilities. But this technological intelligence is being misused by irresponsible people. Fraud mode that utilises artificial intelligence (AI) has recently emerged. In his research, Caldwell explains that AI-powered crime threats are emerging as extensions of existing criminal activity, while others may be new forms.(Caldwell et al., 2020) One example is the case in 2019, where a fraudulent act used artificial intelligence to imitate the voice of the head of a holding company in Germany and contact the CEO of an energy company in the UK, who asked him to send funds totalling €220,000 to a supplier in Hungary.(Schick, 2020) A similar incident occurred in early 2020 when a bank manager in Hong Kong received a call from a person claiming to be a company director. The director said that his company will carry out the acquisition process, requiring assistance authorising transfers of up to

$35 million. Finally, it was discovered that the bank manager had experienced a "deep voice" fraud by the perpetrators who succeeded in cloning the director's voice. Not only is that, but even now, voice cloning technology openly present. Several startup companies, such as London's Aflorithmic, Ukraine's Respeecher and Canada's Resemble.

AI Is running the program.(Brewster, 2021) In subsequent developments, voice cloning finally took place with video cloning which was then known as deepfake.

Deepfake eventually became the research focus by researchers and news in various media. However, the majority of the news that attended often focused on the problems and consequences of deepfakes in four areas, namely (1) focusing on how easy it is to create and share fake video content using deepfakes and applications such as FakeApp; (2) the possibility that deepfakes may lead viewers to false beliefs; (3) deepfakes can undermine the political process in the form of disinformation and fake news; and (4) the use of deepfakes to make non-consensual sexual videos.(Gosse & Burkell, 2020) This paper will discuss the use of deep fakes related to criminal acts in the form of social engineering attacks aimed at obtaining confidential data or extortion by taking into account the current conditions in Indonesia. Therefore, we will describe some phenomena of using deepfake technology and provide a critical view of the urgency of regulation or abuse.

In June 2022, the FBI warned about the threat of information theft using deepfake techniques through job application mode. Applicants use manipulated videos, images or recordings to look and sound like someone else and use other people's personally identifiable information to apply for jobs in IT, programming, database and software companies.(FBI, 2022) Although technology often comes with benefits and potential threats, especially from the perspective of social construction, technology theory views users as agents of technological change and determines the trajectory of technological development. So, it is always necessary to study the benefits of technology that can be developed to overcome the negative aspects that will be present.(Kwok & Koh, 2021)

In general, the term deepfake refers to any digital fake content created with AI techniques, and it became popular after a Reddit user named "deepfakes" claimed in late 2017 for having developed a machine learning algorithm that helps him convert celebrity faces into pornographic videos.(Rathgeb et al., 2022) Basically, deepfake technology relies on machine learning algorithms, which allow software applications to graft faces that are so convincing, because a person's expression is carefully superimposed onto another person's head or an existing recording of a person's movements or mouth sounds, can be used to reverse engineer his or her speech. to pronounce the prepared sentence.(Vaccari & Chadwick, 2020) Other developments have even succeeded in creating new technology that can edit the words spoken by someone just by entering the

text of a new sentence. Not only changing the words, but the technology also allows the deletion of certain words. The engineered video has succeeded in making 58% of respondents believe that the video is genuine.(Fried et al., 2019) The technology is even known to have been misappropriated by selling software that can produce fake text content or vote manipulation and even automate the creation of fake accounts.(Whyte, 2020)

Deepfakes, which some have dubbed the phantom menace in the media, because of the potential problems that can arise considering deepfakes are not limited to visual content but allow voice exchange using audio that can be synthesised by converting or imitating other people's voices, including converting text into sound. specified individual.(Whittaker et al., 2021) Even though identity protective cognition theory states that everyone will always be selective in judging information or news, whether offline or online,(Liv & Greenbaum, 2020) deepfakes still produce information pollution (Hameleers et al., 2022) and can be very dangerous because humans have a strong tendency to trust their own eyes and ears, especially visual information. affect a person's perception of his perceived credibility.(Wobbrock et al., 2021) Thus the proliferation of deepfakes has created new challenges to the trustworthiness of visual experiences and even created negative consequences such as non-consensual pornography, financial fraud and political disinformation.(Johnson & Diakopoulos, 2021) Such as the deepfake video of the Ukrainian President asking Ukrainian soldiers to dispose of their weapons (Dewi, 2022) or even the alleged deepfake sex video of the Malaysian economy minister, Azmin Ali, which in 2019 was spread online.(Kikerpill, 2020) Some parties finally equate the spread of this deepfake 'a zombie horde', which will be difficult to detect and cause division and civil unrest.(Taylor, 2021)

Deepfakes can be abused to exploit someone's implicit trust through video, for example, by impersonating children to older adults via video calls to gain access to their funds, so audio/video impersonation is ranked as the most worrying crime, even if prevention efforts can. This is done through an impersonation detection system. However, this cannot be done in the long term, given that the potential for other new variants of crime that use deepfakes is so high.(Caldwell et al., 2020) In mid-2022, several European politicians even received fake video calls using deepfake technology claiming to be the mayor of Kyiv, Vitali Klitschko. Victims of the video call include the mayor of Berlin, the mayor of Madrid and the mayor of Vienna. In the conversation, the perpetrator

discussed supporting the Ukrainian military because it is currently at war.(Oltermann, 2022) Another case of deepfake abuse occurred in Pennsylvania (Lakhan, 2021). A mother carried out it through photo and video manipulation of 3 cheerleading groups to get rid of her daughter's rival. In the video, the victim is engineered as a bad girl drinking liquor and naked. In the end, Raffaela Spone, the perpetrator of the deepfake video, was arrested in March 2021. However, the police revealed that there was no strong evidence against the fabrication of the video.(Sheets, 2021)

# THEORETICAL FRAMEWORK

The rise of digitalization in various fields of human life is common, where artificial intelligence (AI), as part of the information technology system, has surpassed its role and impact on society (Valenzuela-Fernández et al., 2023). In the era of advanced digitalization, there are concerns about false information and social engineering attacks, which are currently a major concern in the field of information security. Information security relies on user recognition and awareness of the risks associated with social engineering attacks. Unlike typical information threats, social engineering not only exploits technological aspects in the unauthorized collection of information, but also emphasizes the manipulation of human factors to achieve desired results. (Greavu-Şerban & Constantin, 2022)

Research conducted in phishing simulations reports that people with educational and/or professional backgrounds in information technology often perform poorly in social engineering simulations. In (Sütterlin et al., 2022) research, proposed methods and metrics to detect overconfident individuals in terms of deep fake recognition. The proposed overconfidence score flags individuals who overestimate their performance and thus pose a previously unconsidered cybersecurity risk. In this study, and in line with similar research from phishing simulations, individuals with IT backgrounds are particularly prone to overconfidence. So a data-driven approach to identifying at-risk individuals allows educators to provide more targeted education, generate insights into their own assessment shortcomings, and help avoid the self-selection bias that is common with voluntary participation.

Deepfakes are used to spread misinformation, commit fraud, and blackmail innocent people. Deepfakes technology is constantly evolving and attackers now have the ability to create deepfakes in real-time. This poses a significant threat to society as

attackers are beginning to exploit this technology to conduct social engineering attacks.(Frankovits & Mirsky, 2023)

# METHOD

The research method used in this research is a qualitative method, which according to (Sugiyono, 2019), qualitative research methods are often called naturalistic research methods because the investigation is conducted in natural conditions. Activities in data analysis are data reduction, data presentation, and conclusion drawing or verification.

# RESULTS AND DISCUSSION

* 1. DEEPFAKES AND SOCIAL ENGINEERING ATTACKS

It is undeniable that although technological developments are always rooted in good intentions, they still have unintended consequences and are potentially misused for dangerous purposes. In the 1700s, after Ben Franklin created the postal service in the United States, criminals discovered mail fraud. In the 1800s, with the telegraph and telephone, criminals discovered wire fraud. In the twentieth century, when technology invented the internet, it was inevitable that new scams would emerge.(Smith & Browne, 2021) So deepfakes, become one form of the new fraud. Please note that deepfakes are not only misused by creating synthetic videos or images that contain pornographic content. Today, however, Deepfakes pose a significant threat to business activities, where virtual interactions and digital media are the standard forms of communication adopted by most organisations. Deepfake technology has turned to targeting organisations by spreading misinformation and disinformation. Such as spreading misinformation about the CEO of a company who commits fraud and accepts bribes, so that it will damage his reputation and trust in front of shareholders or blackmail scenarios carried out by threatening to release Deepfake videos with harmful content and even creating fictitious figures to steal company secrets.(Sareen, 2022)

Some potential social engineering attacks using deepfake technology can be used in several ways, including: (1) through vishing, through telephone calls and using deepfakes, the recipient of the phone will think that the caller is the person he thinks he is; (2) voice authentication, by using a deepfake, a perfect voice imitation will be produced; (3) blackmail, by lying to have photos/videos that are actually the result of

deepfakes.(Poston, 2022) According to research by Symantec CTO Hugh Thompson, at least three recent attacks took advantage of deepfake voices to defraud companies of millions of dollars, even in one case, up to $10 million. As experienced by Gary Schildhorn, a 67-year-old lawyer, who received a call from someone who claimed to be or at least sounded like his son. Gary revealed that the voice, rhythm and words he used resembled his son. The crying voice on the phone explained that her son had an accident and needed $9,000 to pay the public defender. Within 10 minutes, Schildhorn received another call from someone claiming to be his son's lawyer. Schildhorn was close to reaching his bank to order payment, but before that, he called his daughter-in-law and finally received a call from his son telling him he was fine.(Kostadinov, 2022) Another potential attack, can occur in the zoom application revealed by Matthew Canham, through the example of a video trial where there is a cat acting like a Texas lawyer, with perfect body movements and lips.(Fogerlog, 2022)

At the end of 2017, the news was buzzing about a woman named Vickie O'Shea- Fowler, a CEO of Data Smart Consulting from North Carolina. Vickie appeared on the LinkedIn networking site and requested connections to many key people and security and technology experts. But in the end, it was discovered that Vickie was an unreal figure and a profile created by utilising advanced technology, especially deepfakes. This situation presents the concern of many parties, especially people who receive requests to be connected. At first glance, people can easily think that Vickie is an expert with a good network of friends, as seen from Vickie's connections with key security people such as Chief Information Security Officer 7-Eleven, Director of technology for AT&T, Head of Security. Architecture at Visa Europe and others.(Tomlinson, 2021b)

It is conceivable if ordinary people then judged that Vickie was a real person and then received the malicious emails and links provided by Vickie, the potential for attacks and data leaks was very vulnerable. Even a warning has been issued by the FBI regarding the profile of someone who uses synthetic images with malicious purposes.(Tomlinson, 2021a) Vickie is not an imaginary figure alone. Katie Jones is another example with a fictional persona as a 30-year-old red-haired woman who works for the think tanks of the Brookings Institute. On her LinkedIn profile, Katie is connected to the deputy assistant secretary of state, economist Paul Winfree. Katie claims to have earned a degree in Russian studies from the University of Michigan.(Satter, 2019) Until finally, the initial investigation conducted by Munira Mustaffa revealed that Katie Jones' profile was fake,

and the photo she used was an AI product. In fact, it is suspected that Katie's profile was intended for intelligence purposes.(Hani, 2019) Espionage efforts carried out by Katie Jones on the network platform, finally made the French and German governments issue warnings not to be approached on LinkedIn, even the British Intelligence Agency, MI5 also warned its employees about the potential for social engineering attacks that utilise AI technology.(Filimowicz, 2022)

Vickie and Katie's incident allegedly came with several specific purposes, including: (1) launching a social engineering attack in the form of spearfishing, because the account appears to be a trusted person to then communicate with certain targets to send malware or spyware via email or other messaging systems; (2) gain online and physical access; (3) network mapping, by collecting information about who is connected in certain policy areas.(Giles et al., 2019) Martina Dove explains that it is easier for a cheater if he has gained social trust to eventually have social influence. One of the efforts to gain social influence can not only be done by creating fake profiles but can also buy fake followers to gain more public trust. Fraudsters will be affiliated with groups and use certain commonalities such as areas of expertise, work experience to gain the trust of the group thereby making compliance easier and garnering more potential victims.(Dove, 2020)

Prevention efforts against deepfakes have been around for a long time and are still being carried out today. Several sites have attempted to provide examples of photos of someone using technology, such as https://thispersondoesnotexist.com/ which was created by Philip Wang by developing Nvidia's research using an algorithm called StyleGAN and releasing synthesised photos every 2-3 seconds on the site.(Vincent, 2019) Another site that provides information about photos of technological engineering is https://whichfaceisreal.com. The site is one step for someone to check against questionable personas in cyberspace. Not only happened on the LinkedIn site, Facebook in 2020 deleted 13 accounts for using AI-generated photos with the aim of producing news that cornered the United States government. The account profile informs his profession as a journalist and editor in a news media.(Nimmo et al., 2020)

* 1. DEEPFAKES IN INDONESIA AND REGULATIONS AGAINST THEM

As developments have occurred in the world, deepfake technology has actually been used in Indonesia, considering that deepfakes are not only for destructive purposes, but can be used for constructive purposes such as filmmaking and virtual reality applications,(Shahzad et al., 2022) as used by the Noah band in Indonesia in the video clip process. the music he made.(Azizah, 2022) However, it is undeniable that the use of other deepfakes also comes with negative impacts that are feared to cause, such as political provocation, pornography and social engineering attacks.

The virtual world of Indonesia was shocked at the end of 2020, when hundreds of foreign language accounts supporting Papua's special autonomy appeared on social media. These accounts use Dutch, German, English and Indonesian languages using photos of people that were apparently generated by AI. Initially the number of these accounts was 450 accounts, but in the investigation carried out there were only 100 accounts left. These accounts have been considered to be able to provoke the Papuan people by many parties, because they are campaigning for the independence of West Papua.(Wijaya, 2020) Other fraudulent actions using fake profiles are currently rife with video call sex (VCS) mode. One of the blackmails was carried out by the perpetrator, Iwan Saputra, by creating a fake Facebook account and using a profile photo of a police member with the name Herlan Pratama. After getting to know each other, the perpetrators asked for the victim's whatsapp number to perform VCS, and then recorded the VCS screen. With the threat of spreading the image, the perpetrator asked for Rp. 150 million from the victim.(Review, 2021) Not only that, 12 women in Bandung also became victims of extortion, after the perpetrator, a man who created a fake Facebook profile as a woman, announced job vacancies and asked his victims who registered to send full body videos under the pretext of having an HIV/AIDS free medical test. The perpetrator then extorted Rp.1 million from the victim with the threat of spreading the video.(Iqbal, 2022)

The act of making fake profiles using other people's photos will be increasingly difficult to detect when the perpetrator uses AI to produce synthetic photos, making it difficult for victims to doubt it. In studies of social engineering attacks, this action is often termed “catfishing”, a common method sextortionists use to trick the victim into submitting sexual material or secretly filming the victim while performing a sexual act. Victims are always targeted through fake profiles on social media sites like Facebook or dating apps Tinder and OkCupid. Such was the case of Christopher Patrick Gunn, who

targeted underage girls using a fake Facebook profile in which he pretended to be the "new kid in town," befriended them, and convinced them to send him nude photos. Gunn co-posed as popular singer Justin Bieber on Omegle, a web-based video chat platform, and tricked young fans into sending him nude photos by promising free concert tickets or backstage passes.(Carlton, 2019) Another blackmail was experienced by actress Bella Thorne, whose case began with the threat of distributing Thorne's nude photos. But in the end, Thorne actually resisted by first publishing nude photos of herself, and ended up with a revenge attempt by the perpetrators who made a fabricated video produced from Thorne's recording of her father crying over her father's death combined with a video of a girl masturbating.(Maddocks, 2020)

A study conducted in 2021, has revealed that 96 percent of all deepfake videos are pornographic and non-consensual, with the main victims being women, especially female celebrities whose faces are swapped with explicit pornographic photos/videos. In the study, the deepfake videos had a hundred million views(Giansiracusa, 2021). But not only for extortion, perpetrators often use deepfakes to gain profits by selling video processing results based on orders from people who ask to make videos with the profile they specify. As did a 26-year-old Taiwanese YouTuber, who was arrested for making porn videos using deepfake technology. The victims even exceeded 100 people, including artists, celebrities and public figures. The perpetrators sold the video and made a profit of up to NT$ 5 million.(Pan, 2022)

In Indonesia, the case of deepfakes began to be discussed at the end of 2020, after the presence of reports of pornographic videos allegedly by an artist named Gisella Anastasia. Several parties considered that the video that was spread was the result of engineering using deepfake technology, considering this is the second case that Gisella stumbled upon the problem of pornographic videos. However, it was discovered in the end if the video was genuine. The news of the deepfake resurfaced after the circulation of a 61-second pornographic video similar to another Indonesian artist, Nagita Slavina. In early 2022, the video allegedly by Nagita was widely circulated, but in the end the Indonesian Police, through the Cyber Polda Metro Jaya stated that the video was fake and was the result of engineering using Deepfake technology. Meanwhile, the other victim is an Indonesian celebrity named Titan Tyara, whose photos have been manipulated to be naked and spread on social media. Unlike Nagita, Titan Tyara had received a blackmail threat message so that other photos would not be distributed.

The rise of deepfake cases that have occurred globally and in Indonesia has become an important note for the Indonesian state to prepare preventive measures and law enforcement against these actions. Reflecting on the actions of several states that currently regulate crimes committed using deepfake technology, such as Virginia in 2019 which criminalised the distribution of pornographic content using deepfakes with the intent to coerce, harass and intimidate, and California which introduced the rights for a person to be used in pornographic deepfakes with modest complaints, similar to those legalised in late 2020 in New York. Indonesia seems to have to admit that until now it has not specifically regulated crimes committed using deepfake technology.

In general, Indonesia has declared the protection of its citizens for the feeling of security from the threat of fear as stated in the 1945 Constitution of the Republic of Indonesia Article 28G paragraph (1) which reads "Everyone has the right to protection of personal, family, honour, dignity and property under his control, and has the right to a sense of security and protection from the threat of fear to do or not do something which is a human right. This constitutional mandate is the main basis for the legal protection provided by the state to its citizens. If the case attacks individuals by using electronic media, especially on the Internet, then there is further protection regulated in the Electronic Information and Transactions Law (UU ITE). Legal protection for profiles or accounts that use false identities has been regulated in article 35 which states that if any person intentionally and without rights or against the law manipulates, creates, changes, deletes, destroys Electronic Information and/or Electronic Documents with the aim of making information The Electronic and/or Electronic Document is considered as if the data is authentic. Violations against him are subject to a maximum imprisonment of 12 years. However, it becomes a problem when the profile photo is obtained without using someone else's photo, but it is the result of AI that makes it difficult to fulfil the "as if the data is authentic" element. Thus, it requires further proof if the use of the photo is accompanied by false information that is used to create a certain persona.

In contrast to the case, if the material being distributed is the result of deepfake technology and is disseminated with certain contents, then the protection can refer to Article 27 which ensnares the perpetrator if they intentionally and without rights distribute and/or transmit and/or make accessible Electronic Information and/or Electronic Documents that have contents that violate decency, gambling, contain insults and/or defamation and extortion or threats. Not only that, the provisions of Article 35 of

the ITE Law which states "Every person intentionally and without rights or against the law manipulates, creates, changes, deletes, destroys Electronic Information and/or Electronic Documents with the aim that the Electronic Information and/or Electronic Documents considered as if the data were authentic,” with a criminal sanction of a maximum of 12 years and/or a fine of a maximum of Rp. 12 billion.

If the electronic data contains pornography, then the provisions of Article 4 paragraph (1) of the Pornography Law which prohibits anyone from producing, making, reproducing, duplicating, distributing, broadcasting, importing, exporting, offering, trading, renting, or providing pornography which is explicit punishment can be imposed on the perpetrator with a maximum penalty of 12 years in prison. Meanwhile, if deepfake technology is used by the perpetrator by manipulating a person's voice and admitting that he or she is that person, then the provisions of Article 378 of the Criminal Code which regulates the use of a false name or dignity, with deception or a series of lies to move other people to hand over something to him, are threatened with imprisonment for a maximum of four years. If the fraud is carried out in an electronic transaction, it will be subject to a maximum imprisonment of six years, as regulated in Article 28 paragraph (1) jo. Article 4 paragraph (1) of the ITE Law.

If the regulations that have been enacted in Virginia, California and New York have previously attempted to regulate the impact of deepfake technology related to pornography, regulations in Indonesia actually provide the same protection for its citizens, it can even be applied to the use of fake photos or videos that are intended to look like - processing authentic data. However, the regulation seems to need to be complemented with more specific protection, as has been attempted by the United States, by trying to formulate The Malicious Deep Fake Prohibition Act in 2018, which explicitly mentions deepfake as 'an audiovisual record created or altered in a manner. that the record would falsely appear to a reasonable observer to be an authentic record of the actual speech or conduct of an individual, and the proposed violation is in the form of a maximum prison sentence of 2 years. Indonesia can consider regulations that specifically regulate deepfakes such as the one proposed in the United States by taking into account the birth of a policy regarding the obligation of media companies to implement content moderation. This has been implemented by companies such as, tumblr which since September 2018 has banned deepfakes and Reddit which uses auto moderators to limit abuse of deepfakes.

# CONCLUSION

The misuse of deepfake technology that has been rampant in the realm of politics, pornography to economic losses in the world, has also been felt and happened in Indonesia. Cases of social engineering attacks using phishing, catfishing, and sexortion in Indonesia have been known to use deepfake technology. Law enforcement in Indonesia has actually been regulated in the Criminal Code, the Pornography Law, and the Electronic Information and Transaction Law. However, in order to complement its protection and law enforcement, Indonesia can consider the regulations that are being proposed by the United States by specifically regulating deepfakes and also consider media companies to apply content moderation to deepfake material on its platform.

# ACKNOWLEDGMENTS

Thank you to the Ministry of Education, Culture, Research and Technology of the Republic of Indonesia for funding the research team of the author through the Higher Education Basic Research Grant Program (PDUPT) for the 2022 period, until finally making it the material for writing this paper.

# REFERENCES

Azizah, N. (2022). NOAH Pakai Teknologi Deepfake di Video Klip Kota Mati. https:/[/www.republika.c](http://www.republika.co.id/berita/rhdl2c463/noah-pakai-teknologi-deep-fake-di-video-)o[.id/berita/rhdl2c463/noah-pakai-teknologi-deep-fake-di-video-](http://www.republika.co.id/berita/rhdl2c463/noah-pakai-teknologi-deep-fake-di-video-) klip-kota-mati

Brewster, T. (2021). Fraudsters cloned company director’s voice in $35 million bank heist, police find. Forbes, October 14th Available at: Https://Www. Forbes. Com/Sites/Thomasbrewster/2021/10/14/Huge-Bank-Fraud-Uses-Deep-Fakevoice-Tech- to-Steal-Millions.

Caldwell, M., Andrews, J. T. A., Tanay, T., & Griffin, L. D. (2020). AI-enabled future crime. Crime Science, 9(1), 1–13.

Carlton, A. (2019). Sextortion: The Hybrid Cyber-Sex Crime. NCJL & Tech., 21, 177. https://scholarship.law.unc.edu/ncjolt

Dewi, I. R. (2022). Apa itu Deepfake yang minta tantara Ukraina buang senjata. CNBC Indonesia. https:/[/www.cnbcindonesia.](http://www.cnbcindonesia.com/tech/20220318170614-37-324021/apa-itu-)c[om/tech/20220318170614-37-324021/apa-itu-](http://www.cnbcindonesia.com/tech/20220318170614-37-324021/apa-itu-) deepfake-yang-minta-tentara-ukraina-buang-senjata,

Dove, M. (2020). The psychology of fraud, persuasion and scam techniques: understanding what makes us vulnerable. Routledge.

FBI. (2022). Alter Number I-062822-PSA.

https:/[/www.ic3.gov/Media/Y2022/P](http://www.ic3.gov/Media/Y2022/PSA220628)S[A220628](http://www.ic3.gov/Media/Y2022/PSA220628)

Filimowicz, M. (2022). Deep Fakes: Algorithms and Society. Routledge.

Fogerlog. (2022). Get Ready for Deepfake Phishing Attacks with zoom. https://phishingtackle.com/articles/get-ready-for-deepfake-phishing-attacks/

Frankovits, G., & Mirsky, Y. (2023). Discussion Paper: The Threat of Real Time Deepfakes. ArXiv Preprint ArXiv:2306.02487.

Fried, O., Tewari, A., Zollhöfer, M., Finkelstein, A., Shechtman, E., Goldman, D. B., Genova, K., Jin, Z., Theobalt, C., & Agrawala, M. (2019). Text-based editing of talking- head video. ACM Transactions on Graphics (TOG), 38(4), 1–14. https://doi.org/https://doi.org/10.1145/3306346.3323028

Giansiracusa, N. (2021). How Algorithms Create and Prevent Fake News. Springer. https://doi.org/https://doi.org/10.1007/978-1-4842-7155-1

Giles, K., Hartmann, K., & Mustaffa, M. (2019). The role of deepfakes in malign influence campaigns. NATO StratCom COE.

Gosse, C., & Burkell, J. (2020). Politics and porn: how news media characterizes problems presented by deepfakes. Critical Studies in Media Communication, 37(5), 497–

511. https://doi.org/https://doi.org/10.1080/15295036.2020.1832697

Greavu-Şerban, V., & Constantin, F. (2022). Aspects of Social Engineering and Its Modalities to Countercharge and Prevent It. Informatica Economica, 26(3).

Hameleers, M., van der Meer, T. G. L. A., & Dobber, T. (2022). You won’t believe what they just said! The effects of political deepfakes embedded as vox populi on social media. Social Media+ Society, 8(3), 20563051221116344.

https://doi.org/https://doi.org/10.1177/20563051221116346

Hani, A. (2019). The Lady who Decoded the deepfake Katie Jones. https://themalaysianreserve.com/2019/07/23/the-lady-who-decoded-the-deepfake-katie- jones/

Iqbal, M. (2022). Hasil Pemerasan Video Bugil 12 Gadis Bandung dipakai Judi Online. https:/[/www.detik.com/jabar/hukum](http://www.detik.com/jabar/hukum-dan-kriminal/d-6024330/hasil-pemerasan-video-)-[dan-kriminal/d-6024330/hasil-pemerasan-video-](http://www.detik.com/jabar/hukum-dan-kriminal/d-6024330/hasil-pemerasan-video-) bugil-12-gadis-bandung-dipakai-judi-online

Johnson, D. G., & Diakopoulos, N. (2021). What to do about deepfakes. Communications of the ACM, 64(3), 33–35. https://doi.org/https://doi.org/10.1145/3447255

Kikerpill, K. (2020). Choose your stars and studs: the rise of deepfake designer porn. Porn Studies, 7(4), 352–356.

Kostadinov, D. (2022). Engineering Voice Impersonating from machine learning.

Kwok, A. O. J., & Koh, S. G. M. (2021). Deepfake: a social construction of technology perspective. Current Issues in Tourism, 24(13), 1798–1802. https://doi.org/https://doi.org/10.1080/13683500.2020.1738357

Lakhan, O. (2021). Cyberbullying: teenage girls’ online experiences of, and challenges to sexual harassment.

Liv, N., & Greenbaum, D. (2020). Deep fakes and memory malleability: False memories in the service of fake news. AJOB Neuroscience, 11(2), 96–104. https://doi.org/https://doi.org/10.1080/21507740.2020.1740351

Maddocks, S. (2020). ‘A Deepfake Porn Plot Intended to Silence Me’: exploring continuities between pornographic and ‘political’deep fakes. Porn Studies, 7(4), 415–423. https://doi.org/https://doi.org/10.1080/23268743.2020.1757499

Nimmo, B., Francois, C., Eib, C. S., & Ronzaud, L. (2020). IRA again: Unlucky thirteen. Graphika, September, 1.

Oltermann, P. (2022). European politicians duped into deepfake video calls with mayor of Kyiv. The Guardian, 25. https:/[/www.thegua](http://www.theguardian.com/world/2022/jun/25/european-)r[dian.com/world/2022/jun/25/european-](http://www.theguardian.com/world/2022/jun/25/european-) leaders-deepfake-video-calls-mayor-of-kyiv-vitali-klitschko

Pan, J. (2022). Huang Jie Pans Court for Verdicy in Deepfake case. https:/[/www.taipeitim](http://www.taipeitimes.com/News/taiwan/archives/2022/07/23/2003782274)e[s.com/News/taiwan/archives/2022/07/23/2003782274](http://www.taipeitimes.com/News/taiwan/archives/2022/07/23/2003782274)

Poston, H. (2022). Deepfake Phishing : Can You Trust thal call form the CEO. https://resources.infosecinstitute.com/topic/deepfake-phishing-can-you-trust-that-call- from-the-ceo/

Rathgeb, C., Tolosana, R., Vera-Rodriguez, R., & Busch, C. (2022). Handbook of digital face manipulation and detection: from DeepFakes to morphing attacks. Springer Nature.

https://doi.org/https://doi.org/10.1007/978-3-030-87664-7

Review, R. (2021). Screenshoot Video call Sex jadi Modal Narapidana dalam Lapas di Lampung Peras Warga Riau. https://riaureview.com/news/detail/10115/hukrim/pekanbaru/screenshoot-video-call- sex-jadi-modal-narapidana-dalam-lapas-di-lampung-peras-warga-riau

Sareen, M. (2022). Threats and Challenges by DeepFake Technology. In DeepFakes (pp. 99–113). CRC Press. https://doi.org/10.1201/9781003231493

Satter, R. (2019). Expert : Spy Used AI-generated face to Connect with Targets. https://apnews.com/article/ap-top-news-artificial-intelligence-social-platforms-think- tanks-politics-bc2f19097a4c4fffaa00de6770b8a60d

Schick, N. (2020). Deepfakes: The coming infocalypse. Hachette UK.

Shahzad, H. F., Rustam, F., Flores, E. S., Luís Vidal Mazón, J., de la Torre Diez, I., & Ashraf, I. (2022). A Review of Image Processing Techniques for Deepfakes. Sensors, 22(12), 4556. https://doi.org/https://doi.org/10.3390/s22124556

Sheets, M. (2021). Cops who accused woman, 50, of creating ‘deepfake’ photo and videos of her daughter’s cheerleading rivals admit they never found evidence of manipulated images- but will still prosecute her for harassment. Dailimail.Co.Uk. https:/[/www.dail](http://www.dailymail.co.uk/news/article-9592117/Cops-accused-woman-creating-)y[mail.co.uk/news/article-9592117/Cops-accused-woman-creating-](http://www.dailymail.co.uk/news/article-9592117/Cops-accused-woman-creating-) deepfake-images-never-evidence.html

Smith, B., & Browne, C. A. (2021). Tools and weapons: The promise and the peril of the digital age. Penguin.

Sugiyono. (2019). Metode Penelitian Kuantitatif, Kualitatif, dan R&D. Alfabeta.

Sütterlin, S., Lugo, R. G., Ask, T. F., Veng, K., Eck, J., Fritschi, J., Özmen, M.-T., Bärreiter, B., & Knox, B. J. (2022). The role of IT background for metacognitive accuracy, confidence and overestimation of deep fake recognition skills. International Conference on Human-Computer Interaction, 103–119. https://doi.org/https://doi.org/10.1007/978-3-031-05457-0\_9

Tad, M. C. S., Mohamed, M. S., Samuel, S. F., & Deepa, M. J. (2023). Artificial Intelligence and Robotics and their Impact on the Performance of the Workforce in the Banking Sector. Revista de Gestão Social e Ambiental, 17(6), e03410–e03410. https://doi.org/https://doi.org/10.24857/rgsa.v17n6-012

Taylor, B. C. (2021). Defending the state from digital Deceit: the reflexive securitization of deepfake. Critical Studies in Media Communication, 38(1), 1–17.

Tomlinson, K. (2021a). Can you Spot the Fake ? https://[www.amperesec.com/news/can-](http://www.amperesec.com/news/can-) you-spot-the-fake

Tomlinson, K. (2021b). Vickie doesn’t exist : Deepfakes in your LinkedIn. https:/[/www.amper](http://www.amperesec.com/news/deepfakes-in-your-linkedin)e[sec.com/news/deepfakes-in-your-linkedin](http://www.amperesec.com/news/deepfakes-in-your-linkedin)

Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact

of synthetic political video on deception, uncertainty, and trust in news. Social Media+ Society, 6(1), 2056305120903408.

https://doi.org/https://doi.org/10.1177/2056305120903408

Valenzuela-Fernández, L. A., Ocaña-Fernández, Y. J., Sánchez, M. A. F., Apaza, Y. C. L., Zubieta-Romero, E., & Uribe-Hernández, Y. C. (2023). Law and Artificial Intelligence: Possibilities and Regulations on the Road to the Consummation of the Digital Verdict. Journal of Law and Sustainable Development, 11(6), e1202–e1202. https://doi.org/https://doi.org/10.55908/sdgs.v11i6.1202

Vincent, J. (2019). ThisPersonDoesNotExist. com uses AI to generate endless fake faces. The Verge, 15.

Whittaker, L., Letheren, K., & Mulcahy, R. (2021). The rise of deepfakes: A conceptual framework and research agenda for marketing. Australasian Marketing Journal, 29(3), 204–214. https://doi.org/https://doi.org/10.1177/1839334921999479

Whyte, C. (2020). Deepfake news: AI-enabled disinformation as a multi-level public policy challenge. Journal of Cyber Policy, 5(2), 199–217.

Wijaya, C. (2020). Papua : Mengapa ratusan akun ‘palsu’ berbahasa Belanda muncul mendukung otsus di tengah penolakan sejumlah warga?

Wobbrock, J. O., Hattatoglu, L., Hsu, A. K., Burger, M. A., & Magee, M. J. (2021). The Goldilocks zone: young adults’ credibility perceptions of online news articles based on visual appearance. New Review of Hypermedia and Multimedia, 27(1–2), 51–96. https://doi.org/https://doi.org/10.1080/13614568.2021.1889690