KHATULISTIWA SMART : JOURNAL OF ARTIFICIAL INTELLIGENCE
e-ISSN : XXXX-XXXX
p-ISSN : XXXX-XXXX

Khatulistiwa Smart : Journal of Artificial Intelligence
Vol. 1 No. 02 (2025) : 41-48
Available online at : https://journal.literasikhatulistiwa.org/index.php/smart

Pancasila-Based AI Ethics: Preventing Digital Manipulation and Deepfakes in Society

Anna Sakila [1] , Maulidiah Yasmin[2] , Nidaurrohmah [3] , Royhan Ammar Sulthon[4]
[1234] Universitas Brawijaya, Malang
annasakila17@student.ub.ac.id[1234]

Doi
Received: 27 Mei 2025 *Accepted: 10 Juni 2025* *Published: 30 Juni 2025*

Abstract *Deepfake is an artificial intelligence technology that is developing very rapidly and has a major impact on all aspects such as the industrial sector, education, and digital communications. In the surge in the use of Deepfake, it is not free from its misuse which can harm many people without discrimination. The absence of regulations that specifically regulate the use of Deepfake and AI technology in Indonesia has caused the legal system to be unable to quickly anticipate technological advances. Education that has not yet instilled digital ethics skills has also caused Indonesia to be a country behind in the readiness of AI regulations and ethics. The many problems and negative impacts that arise, this reality shows that the development of policies and digital ethics approaches have not been able to keep up with the development of AI technology and still need a lot of improvement. Therefore, this study aims to overcome the existing shortcomings by offering an AI ethics approach based on noble values in Indonesia. Pancasila as the foundation of the state includes principles such as humanity, social justice, deliberation, and responsibility, which are very appropriate to face ethical challenges in this digital era. This study uses descriptive qualitative methods and non-participatory observation. This study examines the spread of Deepfake content on digital platforms and offers a Pancasila-based ethical framework to address the challenges of AI use. The findings show that Deepfake content is mostly used for non-consensual pornography, political manipulation, and digital fraud. Low digital literacy and the absence of specific regulations exacerbate social impacts such as polarization and loss of public trust. Pancasila, with its principles of humanity, unity, deliberation, and justice, can be the ethical foundation for strengthening regulation, education, and technology in the development of responsible and inclusive AI.*

Keywords *Ethics; Deepfake; Pancasila; Technology; AI*

1. INTRODUCTION

In today's era, artificial intelligence (AI) technology is rapidly advancing, including Deepfake technology. This technology has brought significant impacts on various aspects of human life, ranging from industry, education, to digital communication. Deepfake is a machine learning–based visual and audio content manipulation technology capable of realistically mimicking a person's face and voice, thus posing serious challenges related to ethics, information security, and public trust in digital media. Since its introduction around 2017, the use of Deepfake technology has surged significantly, both in the entertainment sector and in the spread of disinformation (Ajder et al., 2019; Citron & Chesney, 2019).

The Deeptrace (2019) report indicates that 96% of circulating Deepfake content is used for unethical or inappropriate purposes, such as non-consensual pornography and political manipulation. This phenomenon highlights that the more advanced a technology becomes, the greater the potential for misuse if it is not accompanied by strong ethical principles (Ajder et al., 2019; IEEE, 2019).

In Indonesia, the increasing spread of Deepfake content highlights the urgent need to address this issue. The low level of digital literacy among the public and the rapid dissemination of information through social media increase the likelihood of digital manipulation. Indonesia, as one of the countries with the largest social media user populations in the world, is highly exposed to the spread of false information created with the aid of AI technology. The negative impacts caused by this technology are significant, as seen in cases of manipulated videos ahead of the 2019 general election and fraud involving synthetic voices of public figures, demonstrating how this technology can be misused for interests detrimental to the public. As this technology continues to advance, the situation not only threatens individual reputations but also social and political stability on a broader scale (Shevila Kristiyenda, Faradila, & Basanova, 2025).

Another issue arising from this phenomenon is the absence of specific regulations governing the use of Deepfake and AI technology in Indonesia. The current legal system cannot quickly keep up with the rapid advancements in technology. Meanwhile, the education sector in the country has yet to fully instill digital ethics skills, making the younger generation the most vulnerable group to the spread of misinformation and digital manipulation. Nugroho, Siregar, and Anggraini (2021) stated that Indonesia lags behind other ASEAN countries in terms of regulatory preparedness and AI ethics, particularly those grounded in local values. This reality shows that the development of policies and digital ethics approaches in Indonesia has not kept pace with AI technology advancements and still requires significant improvements to face contemporary challenges.

A review of recent publications indicates that most international artificial intelligence ethical frameworks, such as the one proposed by Floridi et al. (2018), strongly emphasize values like justice, transparency, and responsibility. However, these approaches tend to be general and lack consideration of the specific social and cultural contexts of each country. In Indonesia, the application of technology ethics must be integrated with the values of Pancasila, which reflect the nation’s identity, morality, and philosophy of life. Unfortunately, to date, there are still few academic studies that explicitly incorporate Pancasila principles into AI ethics. This gap in the academic literature needs to be addressed promptly to

formulate more relevant and applicable solutions (Irawan, 2024; Harjanto & Najicha, 2024; Sariputta & Najicha, 2023).

Therefore, this study aims to address existing shortcomings by offering an AI ethics approach based on Indonesia's noble values. Pancasila plays a crucial role in guiding the development of regulations and policies, especially those related to information technology. Its values are integrated into various laws and regulations to ensure alignment with the fundamental principles of the state. The use of technology is also emphasized as a tool to promote and instill Pancasila values in the digital era (Sariputta & Najicha, 2023).

This article aims to explore the potential of Pancasila values as a foundation for designing an AI ethics framework in Indonesia, particularly in the context of preventing the misuse of Deepfake technology. This study is expected to provide both theoretical and practical contributions in creating a digital ethics system that aligns with local culture and context. Additionally, it serves as a guide for policymakers, educators, and the general public to respond to technological developments in a wiser and more responsible manner. The article also paves the way for the development of a Pancasila-based digital ethics curriculum that can be implemented from an early age in educational institutions, aiming to shape a generation better prepared to face the complexities of information technology in the future

2. RESEARCH METHODS

2.1. Research Approach and Data Collection

This study uses a descriptive qualitative approach to gain an in-depth understanding of how Pancasila values can serve as the ethical foundation in the development of artificial intelligence (AI), especially in addressing the challenges of Deepfake technology misuse in Indonesia. A qualitative approach was chosen because the topic is complex and closely related to social values, culture, and public perceptions elements that cannot be measured solely by numbers or statistical data. This approach allows researchers to comprehensively explore the social meaning of AI technology use in society. Additionally, it provides space to observe how Pancasila values are upheld or neglected in everyday digital activities.

2.2. Observation and Literature Study Method

The methods used include literature study and non-participatory observation of the spread of Deepfake content on Indonesian social media. This method was chosen to gain a deeper understanding of how Pancasila values can serve as an ethical reference in AI technology development, while maintaining distance from direct involvement with perpetrators or victims of Deepfake content. This approach is also considered more appropriate given the sensitive nature of the issue and the difficulty of direct access.

The researcher observed Deepfake content circulating on digital platforms such as YouTube, TikTok, Instagram, and Twitter/X. The observation aimed to identify the types of content that appeared, how they were disseminated, and the public's response to them. The observation focused on significant moments such as election periods, national events, or hot political issues. Each observed content was categorized based on the type of manipulation (audio, visual, or both), the context of dissemination, and the

intended message. The researcher also recorded public interactions such as comments, number of reshares, and sentiment as forms of response to the content.

2.3. Supporting Document Study

In addition to observation, this study is also supported by data from various documents. The reviewed documents include scientific journals, reports from organizations such as UNESCO and OECD, regulations from relevant ministries, and books discussing Pancasila and technology ethics. The researcher also referred to previous relevant studies, including those on digital literacy and the role of local culture in technology policy development. News articles, public opinions, and social media monitoring reports were also used to provide a more comprehensive contextual overview.

2.4. Data Analysis Techniques

In analyzing the data, the researcher employed a combination of content analysis and thematic analysis. Content analysis was used to trace patterns and hidden messages in documents and digital content, while thematic analysis helped to group data based on emerging themes during the study. Some of the identified themes included public concerns about the spread of Deepfake, ethics in AI usage, and the role of Pancasila values in regulating the digital ecosystem.

The analysis process involved coding relevant data, organizing it into categories, and then formulating themes that reflect the observed phenomena. Data from various sources were compared and combined to find common threads between social practices and relevant ethical theories. Data validity was strengthened by using multiple sources for comparison and conducting cross-checks between observations and documents.

3. RESULTS AND DISCUSSION

Based on non-participant observation of the spread of Deepfake content on digital platforms such as YouTube, TikTok, Instagram, and Twitter/X, as well as literature analysis, several key findings were obtained:

3.1. Types of Circulating Deepfake Content

a. Political Content: Manipulated videos featuring political figures during the 2024 election campaign emerged as one of the most dominant examples. For instance, the falsified voice or facial appearance of a public figure (such as former Indonesian President Soeharto) was used to spread propaganda during the 2024 election (Purba, 2024).
b. Non-Consensual Pornographic Content: The majority of Deepfake content—approximately 96%—is used to produce adult videos without the consent of the original subjects (Deeptrace, 2019, as cited in Chesney & Citron, 2019).
c. Digital Fraud: Fabricated videos featuring public figures are used to deceive individuals or organizations, often by requesting illegal money transfers. For example, a Deepfake video using the face of Prabowo was employed in a fraud scheme by an individual identified as AMA, who was arrested in Central Lampung in January 2025 (Hukmana, 2025).

Sources:
http://tiktok.com/@garudayaksa13/video/7418382815013702917?q=video%20wajah%20prabowo%20dijadikan%20deepfake%20oleh%20AMA&t=1747913564644

3.2. Public Response and Social Impact

a. Deepfake content often triggers public confusion, political polarization, and a breakdown of trust in digital media.
b. Social media communities tend to spread content without verification, especially if it aligns with their ideological biases.
Sources:
https://vt.tiktok.com/ZShnR3s6r/

3.3. Regulatory and Digital Literacy Gaps

a. Indonesia does not yet have specific regulations to address the misuse of Deepfake and AI.
b. Public digital literacy remains low, making the younger generation vulnerable to becoming victims or spreaders of false content.
Sources:
https://www.tiktok.com/@yogaputra0109/video/7400790072490020101

3.4. The Potential of Pancasila Values in AI Ethics

Ethics can be applied to the use of AI by adhering to moral principles and proper rules. Artificial intelligence (AI) technology must be used responsibly and must not violate others' privacy rights or harm certain individuals or groups. Therefore, it is necessary to ensure that its use does not conflict with existing laws or social norms (Misnawati, 2023).

3.5. Integration of Pancasila as the Ethical Framework for AI

Pancasila, as the foundational philosophy of the state, offers a holistic approach that aligns with Indonesia's socio-cultural context. The following are its applications:

a. The First Principle: Belief in One Supreme God
It emphasizes the moral foundation in technology development. AI must be developed with an awareness of ethical responsibility and avoid creating content that undermines the spiritual values of society.
b. The Second Principle: Just and Civilized Humanity
This principle encourages the protection of human rights in the use of AI, including the rights to privacy and truthful information. For example, Deepfake content that damages a person's reputation or demeans their dignity violates this principle.
c. The Third Principle: The Unity of Indonesia
Using AI to strengthen national unity, not to divide. The spread of false information through Deepfake that incites horizontal conflict or political polarization contradicts the principle of unity.
d. The Fourth Principle: Democracy Guided by the Inner Wisdom of Deliberations/Representatives
Prioritizing public participation in the formulation of technology policies. AI regulations should involve various stakeholders, including academics, human rights advocates, and traditional leaders, through deliberation.
e. The Fifth Principle: Social Justice for All the People of Indonesia

Ensuring that the benefits of AI technology are distributed equitably and do not exacerbate social inequality. For example, AI algorithms must not discriminate against marginalized groups.

3.6. Challenges in Implementing Pancasila Ethics

a. Regulatory Limitations:
Although Pancasila provides normative guidance, its implementation requires specific technical regulations, such as laws specifically addressing AI and Deepfake.

b. Low Digital Literacy:
Public education about the risks of Deepfake and digital ethics needs to be expanded through curricula starting from an early age.

c. Technological Globalization:
Indonesia must balance adopting global technologies with protecting local values. For example, foreign digital platforms operating in Indonesia need to comply with Pancasila-based ethical standards.

3.7. Policy and Education Recommendations

a. Regulation:
- The government needs to draft laws on Digital Responsibility and AI that regulate sanctions for Deepfake misuse.
- Collaborate with ASEAN to develop regional AI ethics standards that reflect local values.

b. Education:
- Develop Pancasila-based digital ethics curricula in schools and universities.
- Launch national campaigns to improve digital literacy among the general public.

c. Technology:
- Invest in transparent and accountable AI-based Deepfake detection tools.
- Establish technology ethics research centers involving multidisciplinary fields.

3.8. Theoretical and Practical Contributions

a. Theoretical:
The Pancasila approach complements global AI ethics frameworks that tend to be abstract. This research opens space for further studies on integrating local values into technology regulation in developing countries.

b. Practical:
The findings of this study can serve as a reference for policymakers, educators, and technology developers to create sustainable and inclusive AI systems.

4. CONCLUSION

This study is limited to descriptive analysis and observation of digital content without involving direct respondents. For further validation, quantitative research or in-depth case studies on Deepfake victims are needed. Pancasila offers a relevant ethical framework to address the challenges posed by Deepfake and AI in Indonesia. By integrating values of humanity, justice, and deliberation, Indonesia can develop technology regulations that are not only responsive but also aligned with national interests. Collaboration among government, academics, and society is essential to create a safe, fair, and sustainable digital ecosystem. Through this approach, the study not only illustrates how Deepfake is misused but also critically examines how Pancasila can serve as an ethical foundation for AI development. This research is expected to contribute both theoretically and practically to creating digital ethics

that reflect the character of the Indonesian nation. By not involving direct informants, this approach also maintains ethical considerations in data collection and reduces the risk of bias.

5. DAFTAR PUSTAKA

Wang, D., Zhang, X., Zhang, Z., & Wang, P. (2020). Understanding security failures of multi-factor authentication schemes for multi-server environments. Computers and Security, 88, 101619. https://doi.org/10.1016/j.cose.2019.101619

Ajder, H., Patrini, G., Cavalli, F., & Cullen, L. (2019). The state of deepfakes: Landscape, threats, and impact. Deeptrace.

Citron, D., & Chesney, R. (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. SSRN. https://ssrn.com/abstract=3213954

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Hagendorff, T. (2020). The ethics of AI ethics: An evaluation of guidelines. Minds and Machines, 30(1), 99–120. https://doi.org/10.1007/s11023-020-09517-8

Harjanto, N. S., & Najicha, F. U. (2024). Pendidikan Pancasila sebagai kerangka etika dalam penggunaan kecerdasan buatan. Jurnal Pendidikan Ilmu Pengetahuan Sosial (JPIPS), 16, 30–37. http://ejournal.upr.ac.id/index.php/JP-IPS

Irawan, I. (2024). Penerapan etika Pancasila dalam pengembangan kecerdasan buatan yang bertanggung jawab di Indonesia. Bit-Tech, 7(2), 263–271. https://doi.org/10.32877/bt.v7i2.1740

IEEE. (2019). Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems (1st ed.). IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems.

Kominfo. (2020). Panduan etika digital: Cerdas dan beretika di dunia digital. Kementerian Komunikasi dan Informatika Republik Indonesia.

Kominfo. (2024). Strategi nasional literasi digital 2024–2027. https://literasidigital.id

Kurniawan, W., Arham, M., & Muluk, H. (2024). Social media's influence on political participation: Insights from a systematic review and meta-analysis in Indonesian psychology. Jurnal Psikologi, 51(3), 236. https://doi.org/10.22146/jpsi.101469

Kusnandar, V. B. (2023). 63,1% pengguna internet Indonesia masih percaya hoaks. Databoks Katadata. https://databoks.katadata.co.id

Monasterio Astobiza, A., Ausín, T., Liedo, B., Toboso, M., Aparicio, M., & López, D. (2022). Ethical governance of AI in the Global South: A human rights approach to responsible use of AI. Proceedings, 136. https://doi.org/10.3390/proceedings2022081136

Nugroho, Y., Siregar, M., & Anggraini, R. (2021). Digital literacy and policy gap in Indonesia:

Between awareness and implementation. Jurnal Ilmu Komunikasi, 19(2), 105–120.

Raharjo, S. T. (2020). Integrating local wisdom and national character education in digital learning. Jurnal Pendidikan Karakter, 10(2), 183–194. https://doi.org/10.21831/jpk.v10i2.33619

Rafif, R., Aulia, S., Iksan, F., Fatih, J., & Antoni, H. (2024). Peran Pancasila di era digital dalam mewujudkan penggunaan teknologi yang bertanggung jawab dan beretika. Indonesian Journal of Economics, 2(1).

Shevila Kristiyenda, Y., 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). https://doi.org/10.59246/aladalah.v2i4

Sariputta, A., & Najicha, F. U. (2023). Pancasila dalam koding: Merangkai kreativitas dan tanggung jawab etis di dunia digital.

Undang-undang (UU) Nomor 19 Tahun 2016 tentang Perubahan Atas Undang-Undang Nomor 11 Tahun 2008 Tentang Informasi Dan Transaksi Elektronik. Database Peraturan BPK (2016).

Badan Pusat Statistik. (2022). Statistik penggunaan teknologi informasi dan komunikasi oleh rumah tangga. https://www.bps.go.id

Winoto, Y., Hanifatunnisa, N. A., & Khadijah, U. L. S. (2024). Pemetaan bibliometrik mengenai penelitian warisan budaya digital. Jurnal Artefak, 11(2), 255. https://doi.org/10.25157/ja.v11i2.16460