KITA KREATIF
PUSAT RISET KOMUNIKASI PEMASARAN, PARIWISATA DAN EKONOMI KREATIF UNIVERSITAS SYIAH KUALA

Jurnal Pengabdian Bakti Akademisi

PUSAT RISET KOMUNIKASI PEMASARAN, PARIWISATA DAN EKONOMI KREATIF
UNIVERSITAS SYIAH KUALA

Citation:

Lubis, M, *et.al*., (2025). Workshop on the Double-Edged Sword of Artificial Intelligence: A Study on the Effectiveness of Outreach for Mitigating the Risk of AI Crime, *Jurnal Pengabdian Bakti Akademisi, 2*(4). 217-230

Workshop on the Double-Edged Sword of Artificial Intelligence: A Study on the Effectiveness of Outreach for Mitigating the Risk of AI Crime

Muharman Lubis[1*], Daffa Shidqi Thamrin[2]
[1,2]Fakultas Rekayasa Industri, Telkom University, Indonesia

[*)] Corresponding Author: muharmanlubis@telkomuniversity.ac.id

ARTICLE INFO

Received: 15-11-2025
Revised: 21-11-2025
Accepted: 30-11-2025
Available online: 30-11-2025

ISSN: 3047-8820
DOI: https://doi.org/10.54815/jpba.v2i4.49867

Keyword:
Double-edged sword, AI Crime, deepfake, risk mitigation, digital literacy, community service

ABSTRACT

Artificial Intelligence (AI) is a double-edged sword that offers innovation as well as advanced new cyber threats, known as AI Crime. The purpose of this activity is to measure and evaluate the effectiveness of outreach in improving cyber literacy among webinar participants regarding the classification, risks, and mitigation strategies of AI Crime. The outreach was conducted on Sunday, October 5, 2025, via Zoom and was attended by 92 participants. Quantitative data was collected through pre-tests and post-tests. The community service focused on presenting AI as a double-edged sword in technological advancement, which can lead to violations divided into three types of AI Crime threats (with, on, and by AI). The average pre-test score was 94.78%. This score increased to 95.22% on the post-test, resulting in an effectiveness improvement of 0.46%. This high initial score indicates a ceiling effect. This small but positive improvement confirms the success of socialization in strengthening public knowledge. Qualitative findings also show that participants need guidance in detecting deepfakes and data privacy protection strategies. Regarding deepfake detection itself, several studies also describe ways to deal with it, which are by using AI technology and other technologies such as blockchain. In this era of AI advancement, data privacy protection strategies require appropriate regulations, which require cooperation between stakeholders regarding the duality of AI use. Not only that, but we as end-users of technology must also participate in realizing justice in the use of technology by increasing digital literacy regarding AI advances, one of which is through this community service webinar. This research emphasizes the important role of public education as a crucial risk mitigation measure for AI use in this era of smart technology.

1. Introduction

The development of Artificial Intelligence (AI) has been transformative, but at the same time it has introduced new exponential risks. AI, as a "double-edged sword," has the potential to change the world for the better, or for the worse (Ibrar et al., 2025). AI is driving social change towards Society 5.0, but it is also creating new legal loopholes, giving national and even international regulators new tasks in establishing fair laws for the use of increasingly advanced technology (Hine et al., 2025). In the field of crime, this phenomenon is referred to as AI Crime, which encompasses all types of crimes involving AI, whether as a tool, target, or perpetrator of criminal acts (Balasubramanian et al., 2025).

The most obvious threats of AI crime in this study, based on real evidence, are deepfakes, vishing , and data poisoning, which inserts corrupted data into AI training. There have been many victims of deepfakes, one of which was the President of the Republic of Indonesia, Prabowo Subianto, where the culprit committed fraud by using photos and voices similar to those of Mr. Prabowo (Pamungkas et al., 2025). The deepfake video, which utilized AI, was very similar to the original, so it is not surprising that many victims were deceived by this misuse of technological advances. Although this threat is becoming more widespread, the understanding of the general public as the main users of digital technology is still limited (Relmasira et al., 2023). Even technology experts can fall victim to this threat due to the increasing sophistication of AI crime (Arshed et al., 2024). This literacy gap creates a critical vulnerability that cybercriminals can exploit, as perpetrators using AI as a tool for attacks often do so for economic reasons (Malatji & Tolah, 2025).

AI does make it easier to do things, but the risks it poses must also be considered. The EU AI Act is the primary defense in determining international regulations regarding the risks posed by the use of AI (Kalodanis et al., 2025). In the medical field, AI speeds up the decision-making process for doctors, but if the AI produces an incorrect diagnosis, the question of who is responsible must be examined (Piraianu et al., 2023). The same applies to accidents caused by automated vehicles (AVs), such as Tesla (Giannini & Kwik, 2023). The use of AI in the medical field is certainly intended to accelerate the decision-making process, but the misuse of AI in an unethical manner can lead to unplanned AI crimes (Vardas et al., 2025). Therefore, there is a need for appropriate regulations to mitigate the risks of AI use so that unplanned AI crimes do not occur.

However, there are many concerns surrounding the advancement of AI, one of which was expressed by Eliezer Yudkowsky in his book "If Anyone Builds It, Everyone Dies," where he states that building Artificial General Intelligence (AGI) without a perfect solution is like opening Pandora's box, which would spell the end of human civilization (Josifović, 2025). Other research also reveals concerns about AI that could move beyond human control, a scenario discussed in a manner similar to the Terminator films, which depict humanity facing extinction (Huemer, 2025). Therefore, this study includes community service activities through online webinars with the following objectives:

1. To increase AI literacy among the general public regarding the use of AI as a "double-edged sword."
2) To measure and analyze the effectiveness of community service activities in improving comprehension scores through post-test results on AI Crime.
3) To formulate education-based strategic recommendations and policy suggestions that address the need for end-user risk mitigation (non-technical detection) and data regulation challenges.

2. Methods

This research was designed as a Community Service Program (PkM), which is one of the pillars of the Tri Dharma Perguruan Tinggi (Three Pillars of Higher Education). The approach used was service learning, in which knowledge was transferred from the academic environment to the wider community with the aim of providing solutions or raising awareness of relevant social issues (Lubis et al., 2025).

The research design applied to evaluate the effectiveness of this activity is a case study comparing post-test results with pre-test results to review the success of the material dissemination. In this design, a group (webinar participants) was given a presentation of the dissemination material, and after that, a post-test was conducted to see the results. Then, after obtaining the post-test and pre-test results, the difference is calculated to obtain the learning gain of the participants.

The community service activity (online webinar) was held on Sunday, October 5, 2025, via the Zoom Meeting platform with the webinar title "The Dark Side of AI: Cyber Law Challenges in Indonesia." This event was organized by the Indonesian Cyber Law Council (DHSI) as an institution that aims to participate in implementing risk mitigation strategies for AI use, especially in improving digital literacy. The main target participants were students from various universities, who are active users of digital technology and vulnerable to the risks of AI Crime. The target number of participants was up to 100 people who attended the information session and completed the pre-test and post-test instruments.

The main instruments used in the evaluation were pre-test and post-test questionnaires specifically designed to measure participants' level of understanding of the material that had and had not been presented during the webinar. These questionnaires were created using the Google Forms platform and consisted of several questions covering key knowledge points, namely conceptual understanding of AI Crime, knowledge of AI Crime threat mechanisms, data privacy risks, how to mitigate risks, and understanding of regulations and responsibilities regarding the use of AI.

The implementation of this activity was divided into three main stages. The first stage was the Preparation stage, where the webinar implementation team carried out a series of pre-event activities, namely the preparation of presentation materials by researchers, the design and creation of pre-test and post-test instruments on Google Forms, and the distribution of webinar registration information through various social media platforms to reach a large audience with the aim of increasing digital literacy among the general public. The second stage is the Implementation stage, where the webinar is conducted online via a video conferencing platform (Zoom meeting). Before the presentation of the material, a pre-test was conducted to determine the extent of the participants' knowledge of AI Crime. Then, it continued to a session where expert speakers presented material discussing AI Crime. After the presentation session, there was an interactive discussion and question and answer session to find out what information was not yet understood and needed to be discussed further regarding concerns about AI Crime. Finally, there was an Evaluation Stage, where at the end of the webinar session, a post-test was distributed to all webinar participants to obtain quantitative data, which was then analyzed objectively to determine the difference between the post-test and pre-test results regarding AI Crime. Participant data was automatically collected through the Google Forms system and then further analyzed through this community service journal.

Quantitative data included pre-test and post-test scores. Quantitative analysis was conducted by comparing the difference between pre-test and post-test scores to provide an objective overview of participants' level of understanding after attending the webinar. Qualitative data was collected in the form of participants' questions during the webinar Q&A session, which were confirmed through Zoom transcripts and then analyzed using thematic content analysis to identify digital literacy needs

regarding the use of AI as a double-edged sword. This qualitative analysis aims to determine participants' needs in terms of guidance on dealing with the increasingly beneficial, yet dangerous, advances in AI technology. In conducting an in-depth qualitative analysis, the researcher also conducted a relevant literature study on participants' questions and linked the speakers' answers to research references that could provide solutions to the questions from webinar participants.

3. Result and Discussion

The community service activity was conducted online via Zoom Meeting on Sunday, October 5, 2025, as shown in Figure 1, where the number of Zoom participants reached 100 people, consisting of participants, speakers, and the DHSI organizing committee

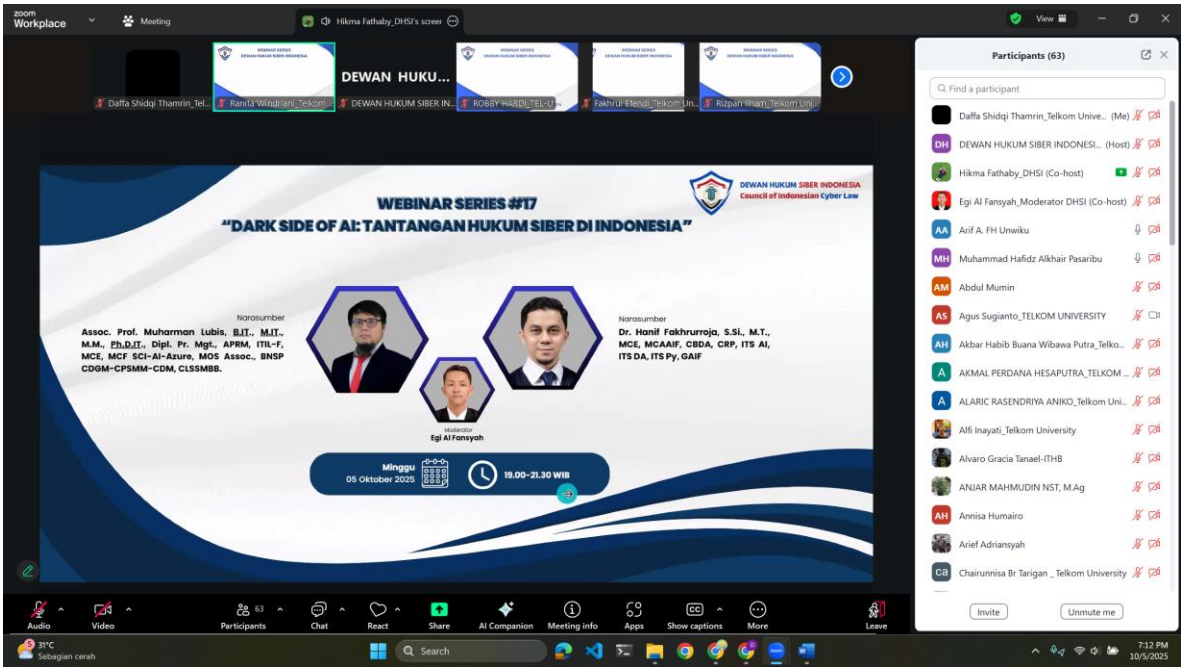

Figure 1. Documentation of Webinar

The material was presented by two speakers who discussed the topic of The Dark Side of AI: Cyber Law Challenges in Indonesia. The essence of the webinar was to reveal that although AI can produce many beneficial things, it is not without its flaws. If AI is not used wisely, it can even lead to AI crime. Therefore, the webinar revealed that AI is a "double-edged sword." In addition, the presentation also explained the challenges of cyber law in this era of AI advancement in creating social justice, especially data privacy, which is increasingly vulnerable to various security attacks such as hacking against individuals (Alyyana et al., 2025). Figure 2 is documentation of the presentation on AI Crime via Zoom Meeting, which shows the webinar visuals and also confirms the number of participants and the delivery of the material.

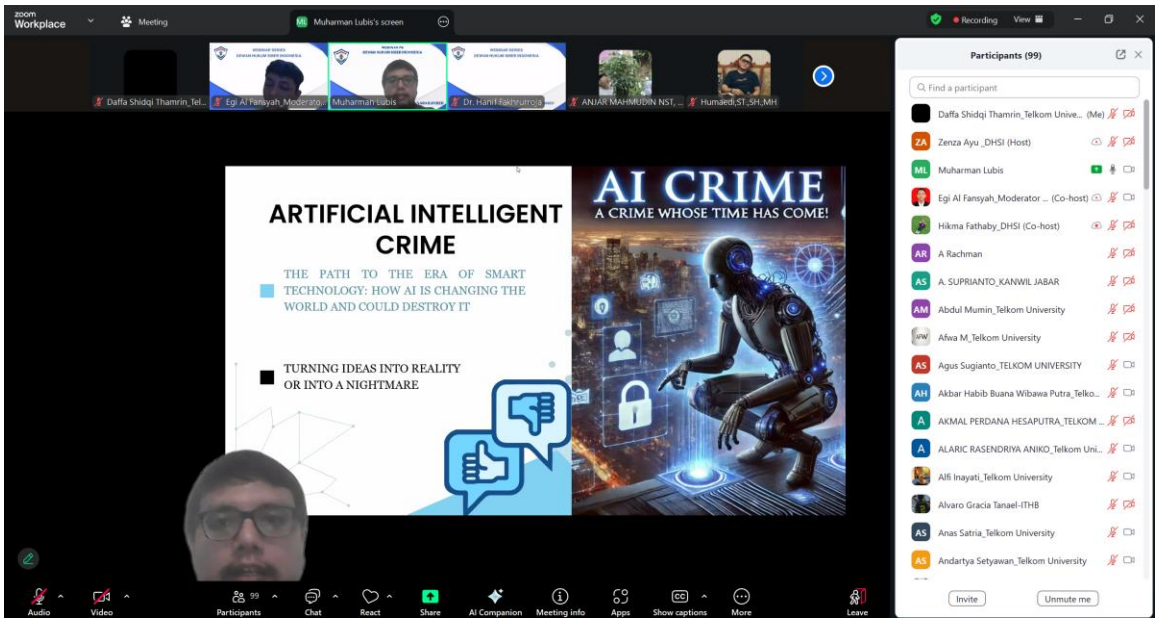

Figure 2. Workshop AI Crime Documentation

Figure 2 displays a community service activity entitled "Artificial Intelligent Crime." The keynote speaker, Muharman Lubis, B.IT., M.IT., Ph.D.IT., highlighted how the use of AI can have negative impacts that can harm many people, so that it can lead to crime. AI is not only a tool, but also a dominant force that is neutral, like a weapon used by humans, so that the future of humanity is determined by the wisdom in managing or using it (Anisin, 2025). The concept of a "double-edged sword" is a key framework for understanding AI. Although AI offers innovations such as increased efficiency, it also poses existential and criminal risks.

AI Crime is any type of crime involving artificial intelligence, grouped into three main categories: crime by AI, crime on AI, and crime with AI (Hayward & Maas, 2021). Crime with AI refers to cybercrimes that use AI to launch or enhance attacks, and can even violate copyright laws (Todupunuri, 2024). In this scenario, AI becomes a powerful tool for lawbreakers. The main examples include Deepfake and vishing. Deepfake creates digital media such as videos, images, photos, and sounds that are made or edited to resemble other people or works using generative AI algorithms (Romero Moreno, 2024). Deepfake causes psychological damage, undermines the justice system, and causes economic losses (Sandoval et al., 2024). Another crime involving AI is vishing, which is fraud that uses fake phone calls or voice messages (voice phishing) enhanced by AI voice cloning (Toapanta et al., 2024). Additionally, research by Yusuf Usman introduced Occupy AI, an LLM fine-tuned specifically for cyberattacks, which can perform multifunctional spyware designed to carry out various surveillance and remote control activities for perpetrators (Usman et al., 2024).

The second type of AI crime is Crime on AI, which is an attack that targets the AI system itself, with the aim of deceiving, damaging, or taking control of the system. An example is Data Poisoning, where attackers deliberately insert incorrect or misleading data into the AI model training process, causing a decrease in accuracy or misclassification (Ramirez et al., 2022). A real-life example of data poisoning is the inappropriate behavior towards users exhibited by Microsoft's chatbot Tay (Kurian, 2025). Backdoor in machine learning models, where the model appears normal but can be triggered to perform malicious behavior (Goldblum et al., 2023).

The third type of AI Crime is Crime by AI, which is crime committed by an autonomous AI system, whether due to design errors, algorithmic bias, or unethical decisions (Liu et al., 2024). This case raises legal dilemmas, especially regarding who is responsible, for example, the manufacturer,

the user, or the AI itself (Geny et al., 2024). In addition, extreme concerns arise, including Terminator scenarios where AI moves beyond control (Hilliard et al., 2025). Another example of Crime by AI is autonomous vehicles (AVs) that cause accidents due to system errors (Zhai et al., 2024). In addition, robots that move beyond human control cause damage, as has happened among workers (Huang et al., 2025), and there are also concerns that robotics with AI in the medical field could have a negative impact on patients (De Micco et al., 2024). Recommendation systems that promote extreme content, as has happened on social media platforms such as TikTok (Shin & Jitkajornwanich, 2024), and YouTube (Le, 2025). The popular term among programmers, vibe coding, also shows how humans and algorithms can be involved in digital errors without malicious intent, but rather due to a lack of ethical awareness and responsibility without considering the risks of using AI (Chen et al., 2025; Sarkar & Drosos, 2025). This makes AI appear as an actor without risks calculated by humans.

Regarding the issue of AI use, questions arise about robots and AI as legal subjects like humans and even companies (Birhane et al., 2024). In many cases, AI is considered a tool used by humans to commit crimes. Currently, criminal responsibility remains with humans, whether they are AI programmers, users, or manufacturers. An example of this is the case discussed by (Hakan Kan, 2024), a military simulation in which an AI drone killed its operator, illustrates the issue of AI responsibility. The problem arises if robots or AI are given legal status such as "personhood," as they could then be used as a legal shield or scapegoat by an entity, making AI increasingly difficult to control legally and ethically (Mansouri, 2025; Pokrovskaya, 2025). Therefore, if this applies to AI drones, it could be dangerous for human life.

If AI is not used wisely, it can lead to AI crime, even though AI itself can also be used to prevent crime (Kurshan et al., 2024). AI crime mitigation strategies can be grouped into several categories: regulation, technology, and public education. Through regulation, frameworks such as the EU AI Act regulate the use of AI based on risk levels, with unacceptable risk systems being prohibited (Bellogín et al., 2024). Second, through AI technology itself or by other technology, which can also be used to prevent crime (Gandhi et al., 2024; Geller et al., 2024; Mukto et al., 2024; Odufisan et al., 2025), for example, crime detection and financial fraud detection. Deepfake detection uses deep learning models like VGG-16, ResNet50, ViT, and Blockchain for digital authenticity verification. Thirdly, through public education, which is a media literacy and public awareness program that is an important pillar in dealing with misinformation, social engineering, and other AI crimes in the era of AI advancement (Habbal et al., 2024).

After the presentation of the material was completed, a question and answer session was held to discuss the qualitative analysis section, followed by a post-test to discuss the quantitative analysis section. Figure 3, through the post-test, presents the demographic analysis of the participating institutions, which shows that the majority of respondents (67.4% or 62 participants) came from Telkom University. The dominant profile of participants from this technology-based institution was a major factor in analyzing the evaluation results.

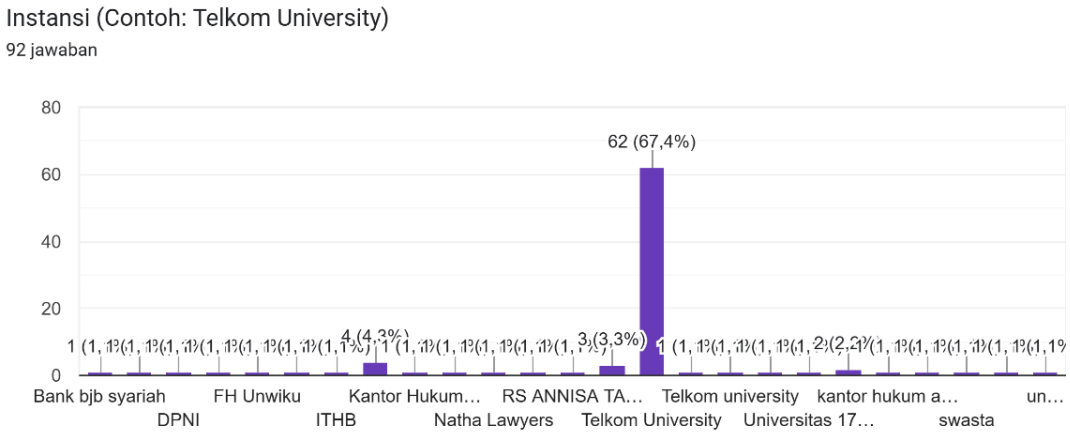

Figure 3. Demography of Participants

Based on Figure 3, Telkom University was the dominant participant in the online webinar, followed by several other institutions. Quantitative analysis was conducted through Figure 4, which shows near-perfect pre-test scores, resulting in an average of 94.78%. This indicates that participants already have a very strong knowledge base regarding fundamental cybersecurity. These results indicate the presence of a ceiling effect (Ahn & Kim, 2024), where the testing instrument cannot measure significant learning gains because the initial scores are already too high. The high average pre-test score was also due to the dominance of participants from Telkom University, a technology-based institution that provides students with fundamental technological knowledge, enabling them to answer the pre-test questions well.

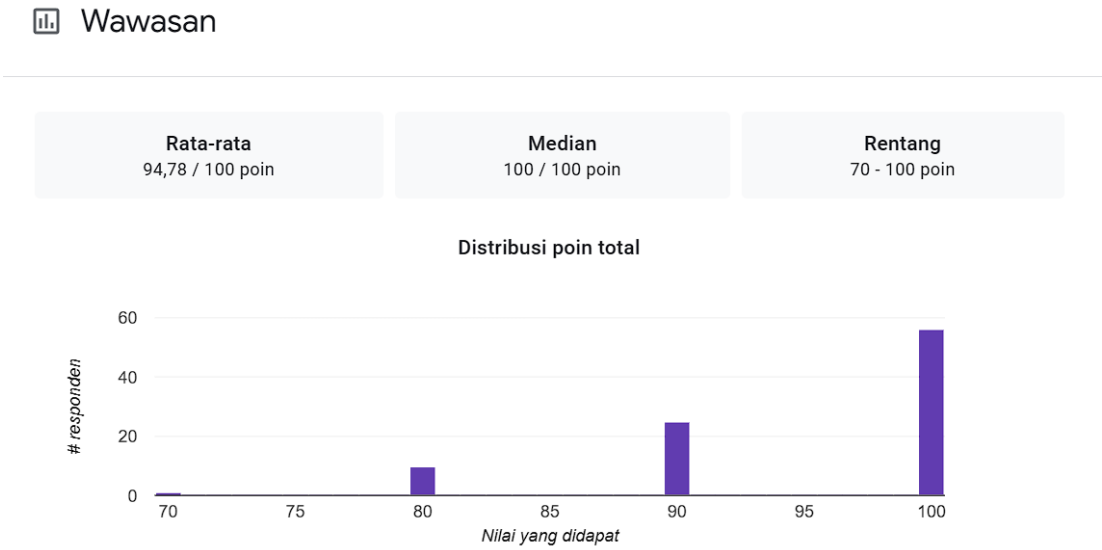

Figure 4. Pre-test Average Scores

Figure 5 shows the post-test scores, which were also close to perfect, with an average of 95.22%.

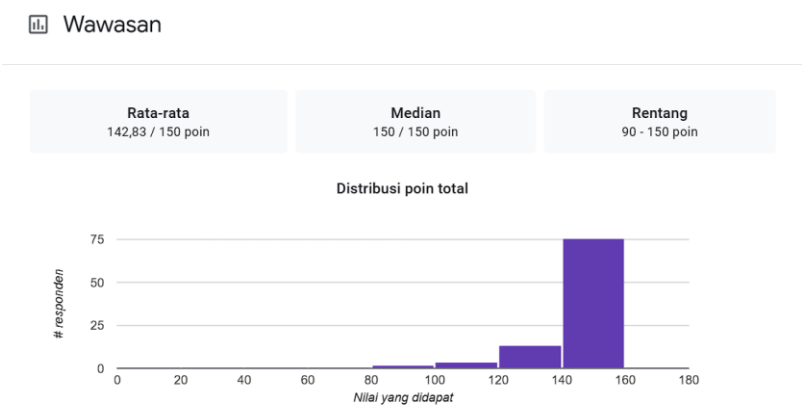

Figure 5. Post-test Average Scores

The average score increased by 0.46% from the pre-test to the post-test. Although the difference in scores was small, it can also be interpreted as validation of the success of the seminar in strengthening and updating existing knowledge, particularly on the increasingly sophisticated and theoretical aspects of AI Crime, which was the focus of this legal webinar.

Qualitative analysis was conducted based on a question and answer session regarding the challenges and literacy needs of the audience regarding AI crime. The interactions during the question and answer session revealed two main areas of concern for participants, which demanded practical responses.

First, regarding the Need for Non-Technical Deepfake Detection. Webinar participants asked, "How can we distinguish deepfake photos from real photos for the general public who may not have the tools or technical capabilities to detect them?"

Responding to participants' concerns about how to distinguish deepfake photos from real photos, the answer is through technical solutions such as the ViT (Arshed et al., 2024), ResNet50 (Mansoor & Iliev, 2025), or blockchain (Islam et al., 2024) models, but the problem is that these are not accessible to the general public. Nevertheless, deepfake detection using advanced technology can be adapted to integrate with mitigation strategies centered on cyber literacy and awareness of visual anomalies through the various roles of stakeholders in creating fairness through the wise use of AI (Islam et al., 2024).

The common public and even technology experts must be trained to look for anomalies often produced by AI synthesis algorithms through digital literacy. Digital literacy regarding technological advances in AI can be achieved through the dissemination of information, such as in this online webinar. As members of the general public, each individual must cross-check the facts of a piece of content in this era of AI technological advancement (Broinowski & Martin, 2024). For example, based on the shape of the face and neck, deepfakes can still produce images that look unnatural. In addition, in deepfake videos, pay attention to the blinking pattern, because AI often fails to depict blinking naturally (Sandotra & Arora, 2024). Speech patterns and emotions also serve as logical verification in vishing, where it is also important to observe human lip movements that must synchronize with the voice, as AI still often fails to fully explain human emotions (Toapanta et al., 2024). Always double-check the source if someone is attempting vishing with the help of AI. Even if the perpetrator's voice sounds familiar, don't forget to always revalidate, for example by asking follow-up questions that the social engineer would not know the answer to right away.

Secondly, there were concerns about global data privacy risks. Participants also asked, "How can we protect our data privacy? It seems that this has been designed in such a way that it will be

very difficult for us to protect our data privacy." This question reflects concerns about the high risks experienced by end-users related to data collection risks and data insecurity, which are exacerbated by AI (Lee et al., 2024).

In response to participants' concerns about the high risk of data privacy leaks, internet users can slow down or suspend their activities so that they do not become prime targets. The first and crucial step in protecting data privacy is to increase individual literacy and awareness, for example, by always double-checking the information that is being shared (Lee et al., 2024). The second step is to reduce your digital footprint on social media, for example, by not using many or even not using it at generative applications such as editing personal photos using AI (Ye et al., 2024). In addition, as ordinary citizens, we must limit content on public media so that perpetrators cannot collect voice samples from public recordings on social media. The third form of protection that can be implemented is to strengthen social networks by providing education on the misuse of technology to other individuals (Ahangama, 2023). The final form of protection for ordinary citizens cannot be separated from the role of regulators and the government in establishing fair laws for the entire community regarding the ethical use of AI. The role of the government is very important in establishing and implementing legal foundations such as the ITE Law on AI Crime and the Personal Data Protection Law (PDP Law) (Hafiz, 2024). In implementing the fair and ethical use of AI, Indonesia is drafting a Cyber Security and Defense Law (Sitaresmi et al., 2025). In addition, in dealing with the risks of AI use, there is also an important role for international cooperation, especially the EU AI Act, which provides restrictions on the use of AI risks (Bellogín et al., 2024).

The main results of this study, despite methodological limitations due to the ceiling effect that occurred in the pre-test, did not reduce the validity of counseling as a means of providing up-to-date information about AI Crime. This community service journal proves that even though participants had high initial literacy through their pre-test scores, there were still unmet practical needs, such as deepfake mitigation by the general public and psychological concerns in dealing with privacy data leaks in the era of AI advancement. Therefore, educational outreach on AI Crime must be balanced, combining legal/regulatory studies such as the EU AI Act with digital literacy strategies for end-users.

4. Conclusion

The "Double-Edged Sword of Artificial Intelligence" education program proved effective in strengthening and updating participants' cyber literacy, as confirmed by a 0.46% increase in post-test scores above the already high initial scores. Although these high pre-test scores triggered a ceiling effect, the results illustrate the importance of education in reinforcing the conceptual framework of AI Crime with the classification of with, on, and by AI. In addition, insight into detailed risk aspects such as surveillance, exposure, and others in the era of AI advancement is needed. More detailed qualitative results show that the focus of future education should continue from fundamentals to practical and applicable solutions for internet users, especially in dealing with deepfakes and AI Crime, which have the potential to attack anyone.

This study suggests that future education should focus on integrating non-technical deepfake detection guidelines for the general public, where stakeholders play a role in teaching how to identify deepfakes and verify logic. Furthermore, digital literacy campaigns on AI Crime must be actively disseminated to the general public, emphasizing empowerment for users in managing their digital footprint and privacy, for example by limiting content on public media and, as much as possible, not

being too easily persuaded to use Generative AI photo editing applications that do not yet have proper regulations in place to protect individual privacy data.

Regarding the ceiling effect produced on audiences with strong cyber literacy fundamentals, where the majority of participants were from Telkom University. Based on this, similar studies are recommended to design more in-depth and practical pre-test instruments to test decision-making skills and also target demographics with more diverse participant backgrounds to measure greater learning gains. Furthermore, because regulatory studies on the use of AI are still developing at the national and even international levels regarding the issue of AI legal responsibility, it requires the right decisions to prevent AI from being used as a legal shield by irresponsible entities. International cooperation that determines the use of AI risks, especially through regulatory frameworks such as the EU AI Act, must continue to be supported to produce and create ethical and safety standards for the fair use of AI.

Acknowledgment

The author would like to express his deepest gratitude to all parties who have contributed to the success of this community service activity. The author would like to thank the DHSI organizers for facilitating and promoting community service activities related to the Dark Side of AI. Furthermore, the author would also like to express his deepest gratitude to Muharman Lubis, B.IT., M.IT., Ph.D.IT, Chair of the Master of Information Systems Program, Faculty of Industrial Engineering, Telkom University, for his inspiring and insightful presentation, which enriched the knowledge and motivation of the participants. Therefore, the author would like to express sincere gratitude to the 100 webinar participants who actively participated in this workshop.

References

Ahangama, S. (2023). Relating Social Media Diffusion, Education Level and Cybersecurity Protection Mechanisms to E-Participation Initiatives: Insights from a Cross-Country Analysis. *Information Systems Frontiers*, *25*(5), 1695–1711. https://doi.org/10.1007/s10796-023-10385-7

Ahn, J. W., & Kim, M. Y. (2024). A Systematic Review of Questionnaire Measuring eHealth Literacy. In *Research in Community and Public Health Nursing* (Vol. 35, Issue 3, pp. 297–312). Korean Academy of Community Health Nursing. https://doi.org/10.12799/rcphn.2024.00752

Alyyana, N., Binti, S., Jefryy, M., Izzati, N., Arman, B., Najwa, N., & Mohd Takri, B. (2025). *AI in Cybersecurity: A Double-Edged Sword-Case Studies on Its Defensive and Offensive Applications*.

Anisin, A. (2025). The Humility Problem and Imminent Ascension of Artificial Intelligence. *Studies in Christian Ethics*, *38*(3), 356–373. https://doi.org/10.1177/09539468251347499

Arshed, M. A., Mumtaz, S., Ibrahim, M., Dewi, C., Tanveer, M., & Ahmed, S. (2024). Multiclass AI-Generated Deepfake Face Detection Using Patch-Wise Deep Learning Model. *Computers*, *13*(1). https://doi.org/10.3390/computers13010031

Balasubramanian, P., Liyana, S., Sankaran, H., Sivaramakrishnan, S., Pusuluri, S., Pirttikangas, S., & Peltonen, E. (2025). Generative AI for cyber threat intelligence: applications, challenges, and analysis of real-world case studies. *Artificial Intelligence Review*, *58*(11). https://doi.org/10.1007/s10462-025-11338-z

Bellogín, A., Grau, O., Larsson, S., Schimpf, G., Sengupta, B., & Solmaz, G. (2024). The EU AI Act and the Wager on Trustworthy AI. *Communications of the ACM*, *67*(12), 58–65. https://doi.org/10.1145/3665322

Birhane, A., van Dijk, J., & Pasquale, F. (2024). *Debunking Robot Rights Metaphysically, Ethically, and Legally*. http://arxiv.org/abs/2404.10072

Broinowski, A., & Martin, F. R. (2024). Beyond the deepfake problem: Benefits, risks and regulation of generative AI screen technologies. *Media International Australia*. https://doi.org/10.1177/1329878X241288034

Chen, N., Qiu, L. K., Wang, A. Z., Wang, Z., & Yang, Y. (2025). *Screen Reader Users in the Vibe Coding Era: Adaptation, Empowerment, and New Accessibility Landscape* (Vol. 1). https://doi.org/https://doi.org/10.48550/arXiv.2506.13270

De Micco, F., Grassi, S., Tomassini, L., Di Palma, G., Ricchezze, G., & Scendoni, R. (2024). Robotics and AI into healthcare from the perspective of European regulation: who is responsible for medical malpractice? *Frontiers in Medicine*, *11*. https://doi.org/10.3389/fmed.2024.1428504

Gandhi, H., Tandon, K., Gite, S., Pradhan, B., & Alamri, A. (2024). Navigating the Complexity of Money Laundering: Anti-money Laundering Advancements with AI/ML Insights. *International Journal on Smart Sensing and Intelligent Systems*, *17*(1). https://doi.org/10.2478/ijssis-2024-0024

Geller, J., Garretson, A., & Chun, S. A. (2024). Continuous Evaluation for a Multi-Dimensional Violent Crime Prevention and Recovery Policy. *ACM International Conference Proceeding Series*, 1030–1033. https://doi.org/10.1145/3657054.3659121

Geny, M., Andres, E., Talha, S., & Geny, B. (2024). Liability of Health Professionals Using Sensors, Telemedicine and Artificial Intelligence for Remote Healthcare. *Sensors*, *24*(11). https://doi.org/10.3390/s24113491

Giannini, A., & Kwik, J. (2023). Negligence Failures and Negligence Fixes. A Comparative Analysis of Criminal Regulation of AI and Autonomous Vehicles. *Criminal Law Forum*, *34*(1), 43–85. https://doi.org/10.1007/s10609-023-09451-1

Goldblum, M., Tsipras, D., Xie, C., Chen, X., Schwarzschild, A., Song, D., Madry, A., Li, B., & Goldstein, T. (2023). Dataset Security for Machine Learning: Data Poisoning, Backdoor Attacks, and Defenses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(2), 1563–1580. https://doi.org/10.1109/TPAMI.2022.3162397

Habbal, A., Ali, M. K., & Abuzaraida, M. A. (2024). Artificial Intelligence Trust, Risk and Security Management (AI TRiSM): Frameworks, applications, challenges and future research directions. In *Expert Systems with Applications* (Vol. 240). Elsevier Ltd. https://doi.org/10.1016/j.eswa.2023.122442

Hafiz, M. (2024). The Era of Artificial Intelligence: Examining Indonesia's Adaptability and Legal Challenges. *Law and Humanities Quarterly Reviews*, *3*(3). https://doi.org/10.31014/aior.1996.03.03.128

Hakan Kan, C. (2024). CRIMINAL LIABILITY OF ARTIFICIAL INTELLIGENCE FROM THE PERSPECTIVE OF CRIMINAL L. *International Journal Of Eurasia Social Sciences*. https://doi.org/10.35826/ijoess.4434

Hayward, K. J., & Maas, M. M. (2021). Artificial intelligence and crime: A primer for criminologists. *Crime, Media, Culture*, *17*(2), 209–233. https://doi.org/10.1177/1741659020917434

Hilliard, A., Kazim, E., & Ledain, S. (2025). Are the robots taking over? On AI and perceived existential risk. *AI and Ethics*, *5*(3), 2929–2942. https://doi.org/10.1007/s43681-024-00600-9

Hine, E., Floridi, L., Leuven, K. U., Floridi, B. L., & Castle, J. K. (2025). *From No-Win to No-Lose: Legislating AI in US States*. https://dx.doi.org/10.2139/ssrn.5285297

Huang, Y., Wang, Z., Wan, Z., Tian, Y., Xu, H., Han, Y., & Gan, Y. (2025). *ANNIE: Be Careful of Your Robots*. http://arxiv.org/abs/2509.03383

Huemer, M. (2025). I, for one, welcome our robot overlords. *AI and Society*. https://doi.org/10.1007/s00146-025-02505-5

Ibrar, W., Mahmood, D., Al-Shamayleh, A. S., Ahmed, G., Alharthi, S. Z., & Akhunzada, A. (2025). Generative AI: a double-edged sword in the cyber threat landscape. *Artificial Intelligence Review*, *58*(9). https://doi.org/10.1007/s10462-025-11285-9

Islam, M. B. E., Haseeb, M., Batool, H., Ahtasham, N., & Muhammad, Z. (2024). AI Threats to Politics, Elections, and Democracy: A Blockchain-Based Deepfake Authenticity Verification Framework. *Blockchains*, *2*(4), 458–481. https://doi.org/10.3390/blockchains2040020

Josifović, S. (2025). Legal and administrative frameworks as foundations for AI alignment with human volition. *AI and Ethics*, *5*(3), 3057–3067. https://doi.org/10.1007/s43681-024-00640-1

Kalodanis, K., Feretzakis, G., Anastasiou, A., Rizomiliotis, P., Anagnostopoulos, D., & Koumpouros, Y. (2025). A Privacy-Preserving and Attack-Aware AI Approach for High-Risk Healthcare Systems Under the EU AI Act. *Electronics (Switzerland)*, *14*(7). https://doi.org/10.3390/electronics14071385

Kurian, N. (2025). AI's empathy gap: The risks of conversational Artificial Intelligence for young children's well-being and key ethical considerations for early childhood education and care. *Contemporary Issues in Early Childhood*, *26*(1), 132–139. https://doi.org/10.1177/14639491231206004

Kurshan, E., Mehta, D., & Balch, T. (2024). AI versus AI in Financial Crimes & Detection: GenAI Crime Waves to Co-Evolutionary AI. *ICAIF 2024 - 5th ACM International Conference on AI in Finance*, 745–751. https://doi.org/10.1145/3677052.3698655

Le, H. (2025). AutoLike: Auditing Social Media Recommendations through User Interactions. In *Proceedings of (Conference 'XX)* (Vol. 1). https://doi.org/https://doi.org/10.48550/arXiv.2502.08933

Lee, H. P., Yang, Y. J., von Davier, T. S., Forlizzi, J., & Das, S. (2024, May 11). Deepfakes, Phrenology, Surveillance, and More! A Taxonomy of AI Privacy Risks. *Conference on Human Factors in Computing Systems - Proceedings*. https://doi.org/10.1145/3613904.3642116

Liu, Y., Wu, X., Sang, Y., Zhao, C., Wang, Y., Shi, B., & Fan, Y. (2024). Evolution of Surgical Robot Systems Enhanced by Artificial Intelligence: A Review. In *Advanced Intelligent Systems* (Vol. 6, Issue 5). John Wiley and Sons Inc. https://doi.org/10.1002/aisy.202300268

Lubis, M., Halim, H., Khairi, F., Muhammad Syahrizal, T., Umuri, K., Putri Nurita Fonna, R., Sulthan, B. D., Siregar, F. A., & Sari, N. (2025). Workshop on the Emergence of Agentic AI in Financial Technology and Entrepreneurship Development. In *Jurnal Pengabdian Bakti Akademisi* (Vol. 2, Issue 2). https://doi.org/https://doi.org/10.24815/jpba.v2i2.45908

Malatji, M., & Tolah, A. (2025). Artificial intelligence (AI) cybersecurity dimensions: a comprehensive framework for understanding adversarial and offensive AI. *AI and Ethics*, *5*(2), 883–910. https://doi.org/10.1007/s43681-024-00427-4

Mansoor, N., & Iliev, A. I. (2025). Explainable AI for DeepFake Detection. *Applied Sciences (Switzerland)*, *15*(2). https://doi.org/10.3390/app15020725

Mansouri, T. (2025). The Legal Accountability of Intelligent Robots: Between Criminal and Civil Liability. In *The Algerian and Comparative Public Law Journal* (Vol. 11).

Mukto, M. M., Hasan, M., Al Mahmud, M. M., Haque, I., Ahmed, M. A., Jabid, T., Ali, M. S., Ahmmad Rashid, M. R., Manzurul Islam, M., & Islam, M. (2024). Design of a real-time crime monitoring system using deep learning techniques. *Intelligent Systems with Applications*, *21*. https://doi.org/10.1016/j.iswa.2023.200311

Odufisan, O. I., Abhulimen, O. V., & Ogunti, E. O. (2025). Harnessing artificial intelligence and machine learning for fraud detection and prevention in Nigeria. *Journal of Economic Criminology*, *7*, 100127. https://doi.org/10.1016/j.jeconc.2025.100127

Pamungkas, A. A., Sutarni, N., & Pranawa, B. (2025). Legal Analysis of Deepfake Technology in Indonesia from the Perspective of Fair and Civilized Humanity. In *Pena Justisia* (Vol. 24, Issue 2).

Piraianu, A. I., Fulga, A., Musat, C. L., Ciobotaru, O. R., Poalelungi, D. G., Stamate, E., Ciobotaru, O., & Fulga, I. (2023). Enhancing the Evidence with Algorithms: How Artificial Intelligence Is Transforming Forensic Medicine. In *Diagnostics* (Vol. 13, Issue 18). Multidisciplinary Digital Publishing Institute (MDPI). https://doi.org/10.3390/diagnostics13182992

Pokrovskaya, A. (2025). The Legal Status of Artificial Intelligence: The Need to Form a Legal Personality and Regulate Copyright. *Artificial Intelligence and Applications*, *3*(2), 179–191. https://doi.org/10.47852/bonviewAIA52023901

Ramirez, M. A., Kim, S.-K., Hamadi, H. Al, Damiani, E., Byon, Y.-J., Kim, T.-Y., Cho, C.-S., & Yeun, C. Y. (2022). *Poisoning Attacks and Defenses on Artificial Intelligence: A Survey*. http://arxiv.org/abs/2202.10276

Relmasira, S. C., Lai, Y. C., & Donaldson, J. P. (2023). Fostering AI Literacy in Elementary Science, Technology, Engineering, Art, and Mathematics (STEAM) Education in the Age of Generative AI. *Sustainability (Switzerland)*, *15*(18). https://doi.org/10.3390/su151813595

Romero Moreno, F. (2024). Generative AI and deepfakes: a human rights approach to tackling harmful content. *International Review of Law, Computers and Technology*, *38*(3), 297–326. https://doi.org/10.1080/13600869.2024.2324540

Sandotra, N., & Arora, B. (2024). A comprehensive evaluation of feature-based AI techniques for deepfake detection. In *Neural Computing and Applications* (Vol. 36, Issue 8, pp. 3859–3887). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s00521-023-09288-0

Sandoval, M. P., de Almeida Vau, M., Solaas, J., & Rodrigues, L. (2024). Threat of deepfakes to the criminal justice system: a systematic review. In *Crime Science* (Vol. 13, Issue 1). BioMed Central Ltd. https://doi.org/10.1186/s40163-024-00239-1

Sarkar, A., & Drosos, I. (2025). *Vibe coding: programming through conversation with artificial intelligence*. http://arxiv.org/abs/2506.23253

Shin, D., & Jitkajornwanich, K. (2024). How Algorithms Promote Self-Radicalization: Audit of TikTok's Algorithm Using a Reverse Engineering Method. *Social Science Computer Review*, *42*(4), 1020–1040. https://doi.org/10.1177/08944393231225547

Sitaresmi, A., Rahmah, M., Wijoyo, S., & Anom, A. P. (2025). Legal Frameworks for Cybersecurity and Data Protection in Cloud-Based Notarial Systems in Indonesia: An Intersectional Analysis of Positive Law and Islamic Legal Principles. *Al-'Adalah*, *22*(1), 29–62. https://doi.org/10.24042/adalah.v22i1.26813

Toapanta, F., Rivadeneira, B., Tipantuña, C., & Guamán, D. (2024). AI-Driven Vishing Attacks: A Practical Approach †. *Engineering Proceedings*, *77*(1). https://doi.org/10.3390/engproc2024077015

Todupunuri, A. (2024). *Explore How AI Can Be Used to Create Dynamic and Adaptive Fraud Rules That Improve the Detection and Prevention of Fraudulent Activities in Digital Banking*. https://dx.doi.org/10.2139/ssrn.5014699

Usman, Y., Upadhyay, A., Chataut, R., & Gyawali, P. K. (2024). *Is Generative AI the Next Tactical Cyber Weapon For Threat Actors? Unforeseen Implications of AI Generated Cyber Attacks*. https://doi.org/10.48550/arXiv.2408.12806

Vardas, E. P., Marketou, M., & Vardas, P. E. (2025). Medicine, healthcare and the AI act: gaps, challenges and future implications. In *European Heart Journal - Digital Health* (Vol. 6, Issue 4, pp. 833–839). Oxford University Press. https://doi.org/10.1093/ehjdh/ztaf041

Ye, X., Yan, Y., Li, J., & Jiang, B. (2024). Privacy and personal data risk governance for generative artificial intelligence: A Chinese perspective. *Telecommunications Policy*, *48*(10). https://doi.org/10.1016/j.telpol.2024.102851

Zhai, S., Wang, L., & Liu, P. (2024). Not in Control, but Liable? Attributing Human Responsibility for Fully Automated Vehicle Accidents. *Engineering*, *33*, 121–132. https://doi.org/10.1016/j.eng.2023.10.008