Applied Science and Innovative Research
ISSN 2474-4972 (Print) ISSN 2474-4980 (Online)
Vol. 10, No. 3, 2026
www.scholink.org/ojs/index.php/asir

Original Paper

An Assessment of Challenges and Opportunities for Forensic Investigators: Capitalizing on Artificial Intelligence and Digital Forensics

Hassan Fayyad-Kazan[1], Jelnar Khattar[2], Hussin Jose Hejase[3], Ale J. Hejase[4]

[1] Computer Science & Engineering Department, Kuwait College of Science & Technology, Kuwait City, Kuwait

[2] Master's Student, Forensic Sciences, Faculty of Sciences I, Lebanese University, Beirut, Lebanon

[3] Basic and Applied Sciences Research Center, Al Maaref University, Beirut, Lebanon; IEEE Senior Member

[4] Principal Researcher, Said NGO, Beirut, Lebanon

Correspondence: Hussin J. Hejase, E-mail: hussin.hejase@mu.edu.lb

Abstract

This paper investigates the potential of Artificial Intelligence (AI) to strengthen and sustain digital forensics capabilities. The aim is to examine applications for detecting AI-generated images, uncovering hidden messages, and identifying phishing URLs. While acknowledging the inherent ethical and practical challenges of AI implementation, the research emphasizes its potential to revolutionize digital forensics by equipping investigators with powerful new tools. The findings show the complexities of training AI models and the key role of ongoing research and development in keeping ahead of evolving cyber threats.

Keywords

Artificial Intelligence, digital forensics, challenges, opportunities, cyber vulnerabilities

1. Introduction

Living in a world in which technology has become an essential part of people's lives, access to Artificial Intelligence (AI) tools has become easier to use and sometimes free, eliciting concerns (Hawking, 2016; Hussein & Hejase, 2022). Some people will inevitably use these tools for malicious purposes. Digital forensics is the process of collecting, analyzing, and presenting digital evidence in court (Franke, Hjelmås, & Wolthusen, 2013), and relies on established methods that regularly struggle to collect and analyze this large volume of data. To increase efficiency and save time, it is necessary to automate some tasks using AI, which has a rich history dating back to the early days of computing

(Muthukrishnan, Maleki, & Ovens, 2020). In recent years, AI has experienced a resurgence of interest due to advances in machine learning and deep learning (Abbas & Hurr, 2023).

This research focuses on three key areas in which AI is beneficial for digital forensic investigators. Firstly, AI-generated images blur the line between real and fabricated (Göring, Ramachandra Rao, Merten, & Raake, 2023), posing concerns surrounding fake news and identity theft. While these realistic creations raise significant ethical questions, they also point out the need for advanced detection methods. Similarly, text steganography, the ability to hide information within seemingly harmless text data (Khan, Ali, & Asghar, 2015; Fayyad-Kazan, Hejase, El Dirani, & Rkein, 2021), makes it difficult to detect potentially malicious content. This challenge underscores the need for robust data analysis techniques that AI can deliver. Finally, URL phishing, a deceptive practice that exploits human trust through misleading messages and links (Zhang, Hamadi, Damiani, et al., 2022; Fayyad-Kazan et al., 2024), often results in the compromise of sensitive information. While user education remains essential in combating phishing incidents, AI-powered solutions can provide an additional layer of protection by identifying and filtering malicious URLs. In response to these arising challenges, AI offers encouraging opportunities for digital forensics.

1.1 Research Questions

1. Detecting AI-generated images: Can AI algorithms effectively distinguish computer-generated visuals from authentic photographs?
2. Uncovering hidden messages: Can AI tools be employed to expose hidden communication embedded within seemingly harmless text data?
3. Identifying phishing URLs: Can AI models precisely differentiate malicious phishing attempts disguised as legitimate websites?

This paper consists of five sections. Section one covers the introduction and research questions. Sections two and three constitute the literature review and methodology, while section four relates to the results and findings. Section five provides a conclusion and recommendations.

2. Literature Review

This literature review explores the growing field of AI-powered digital forensics, illustrating its potential to reinvent investigations while concurrently addressing its limitations and ethical implications.

2.1 Promising Applications of AI

- Automating monotonous tasks: Several studies, such as those by Elsaesser and Tanner (2001) and Mohammed, Clarke, & Li (2016) showcase how AI can automate tedious tasks in data collection, analysis, and evidence processing, freeing investigators for more complex work.
- Enhanced accuracy: AI's ability to examine vast amounts of data is demonstrated in studies by Mitchell (2010) and Krivchenkov, Misnevs, & Pavlyuk (2019). This capacity may lead to more accurate investigations by identifying sophisticated patterns that might escape human detection.

- Detecting emerging threats: Irons and Lallie (2014) posit that AI's potential for identifying new cybercrime tactics empowers investigators to stay ahead of evolving threats and proactively combat cybercrime.

2.2 Specific Applications

- AI-generated image detection: de Rezende, Ruppert, & Carvalho (2017), Arora and Soni (2021), and Bird and Lotfi (2023) demonstrate promising techniques for detecting AI-generated images, important for combating misinformation and privacy concerns.
- Text steganography detection: Oplatková et al. (2008), Yang et al. (2019), and Li, Wang, & Liu (2023) present successful AI models for detecting hidden information embedded within text data.
- URL phishing detection: Studies by Galego Hernandes, Floret, Cardozo De Almeida, et al. (2021), Bouijij and Berqia (2021), and Basit, Zafar, Liu, et al. (2021) showcase progress in AI-powered URL phishing detection, important for safeguarding user privacy and security.

2.3 Challenges and Considerations

- Data volume and complexity: MacDermott, Baker, & Shi (2018) noted that the ever-increasing volume of digital evidence and the elaborate nature of modern networks pose challenges for AI systems.
- Reliability, transparency, and explainability: Studies by Costantini, De Gasperis, & Olivieri (2019) and Solanke (2022) point out the need for reliable, transparent, and explainable AI models in digital forensics. Ensuring human oversight and understanding of AI decision-making processes is fundamental.
- Attribution and bias: Identifying the responsible AI system and its motivations (Schneider and Breitinger, 2020) and guaranteeing unbiased training data (Silberg and Manyika, 2019; Lamberti, 2023) are critical challenges that require ongoing research and development.
- Ethical aspects: The integration of AI raises multifaceted ethical concerns, as highlighted by Stahl, Carroll-Mayer, Elizondo, et al. (2012), Balogun and Zuva (2017), and Alaa El-Din (2022). Striking a balance between effective investigations, personal privacy rights, and ethical development continues paramount. Frameworks such as those proposed by Chinnikatti (2018) and Maratsi et al. (2022) provide valuable guidance for dealing with these ethical issues.

3. Materials and Methods

This paper proposes three research questions to investigate and test the extent to which artificial intelligence (AI) and digital forensics can be leveraged to detect AI-generated images, reveal concealed messages, and identify phishing URLs. Therefore, three experiments were proposed and designed.

3.1 Experiment 1: Examining the Trustworthiness of AI in Detecting AI-Generated Images

This experiment evaluated the effectiveness of three free AI tools for identifying AI-generated images across a varied dataset. We used a sample of 122 images, categorized as authentic or AI-generated, to ensure statistically valid results.

3.1.1 Tools and Data

Three AI detection tools (available for free) were used for the analysis: AI Tool 1 (AI or not), AI Tool 2 (Hugging Face), and AI Tool 3 (Hive Moderation).

Briefly presented

AI or Not: AI detector for deepfakes of all sorts, including images (AI or Not, 2024).

Hugging Face: According to the website Huggingface (2024), "This application serves as a proof-of-concept for utilizing a ViT model to determine whether AI was used to create an artistic image" (para 1). Moreover, the designer blog contends that "AI art is undergoing a surge thanks to the open-source release of Stable Diffusion, a text-to-image tool produced by a team of computer vision researchers funded by Stability.ai. While there have been previous programs that are comparable (such as VQGAN+CLIP, Disco Diffusion, DALLE-2, Midjourney, and Craiyon), Stable Diffusion is the first that is available as an open-source project and is free of cost and can form realistic, highly detailed images in response to cues from the user" (Mathew Maybe, 2022, para 1).

Hive Moderation: This website offers automated content moderation solutions with human-level accuracy. It supports solutions for detecting visual, text, audio, and AI-generated content, and it comes with a moderation dashboard (Hive Moderation, 2024).

The dataset comprised 122 images: 61 genuine images and 61 created by AI tools.

3.1.2 Evaluation Process

Images were analyzed using the three tools. Each tool's response per image was recorded. Also, an analysis of its ability to distinguish AI-generated content was described.

Performance was measured using the following key metrics (Gillham, 2024):

- ❖ Accuracy: Overall correctness of AI detection.
- ❖ Precision: Ability to correctly identify AI-generated content when claimed.
- ❖ Recall: Ability to identify all AI-generated content correctly.
- ❖ F1-Score: Balances precision plus recall for reliable detection.
- ❖ False Positive Rate: The tendency to incorrectly classify genuine images as AI-generated (Gillham, 2023).
- ❖ False Negative Rate: The tendency to miss AI-produced content (Gillham, 2023).

3.1.3 Desired Outcome

This evaluation aimed to assess the reliability and accuracy of these free AI tools for distinguishing AI-generated images, informing their potential and limitations for use in digital forensics.

3.2 Experiment 2: Unveiling Covert Communication: Examining Text Steganography and AI Detection Methodology

The increasing popularity of apps that hide messages in plain text raises concerns about their potential misuse for covert communication. This experiment studies the potential of AI to detect hidden messages through text steganography in WhatsApp chats.

3.2.1 Objective

Develop a method to identify potentially steganographic sentences within WhatsApp chat data.

This method could involve an AI model recognizing unusual characters used for steganography purposes.

3.2.2 Experimental Setup

We analyzed two types of WhatsApp conversations:

(1) Authentic Interaction: Mimicked a real conversation between friends.

(2) Hidden Message Application: Used a designated app (e.g., HideMessage) to embed hidden messages within text.

Both conversations were carefully recorded and exported for analysis.

3.2.3 Method

1. Create two conversation types: Authentic and hidden message application usage.

2. Record and export both conversation types from WhatsApp.

3. Label the exported data as "Hidden" or "Not Hidden." You can access these data via:

drive link (https://drive.google.com/file/d/1LgmsextjmbZf4dSWkvXu_O88Wciee_Gx/view)

Or

Note: Because of limitations in our knowledge of AI model training, we were unable to complete the following steps:

4. Utilize an AI tool to examine the data, searching for specific patterns of characters used for encoding hidden messages.

5. Analyze results to identify these patterns.

Although the researchers lacked the training to complete steps 4 and 5, this experiment demonstrates the potential use case for AI in detecting text steganography in WhatsApp. Further research with expertise in AI model training is necessary to develop and implement a robust detection system.

3.3 Experiment 3: Assessing the Potential of AI in URL Phishing Detection

Due to restricted access to real phishing data, this experiment uses a SWOT analysis to assess the strengths, weaknesses, opportunities, and threats of applying AI to URL-based phishing detection. This analysis intends to examine the benefits and drawbacks of AI in this context, informing future research and development efforts.

4. Results and Findings

4.1 Experiment 1: Detection Results of AI Tools for Identifying AI-Generated Images:

Our experiment examines the effectiveness of three AI tools (AI or Not, Hugging Face, and Hive Moderation) at detecting AI-generated images in a carefully selected dataset. To understand the intricacies of their performance, we analyzed the estimated percentage of AI presence in each image alongside the detection outcomes, categorized as True Positives (accurately identified AI images), True Negatives (correctly identified non-AI images), False Positives (images mistakenly identified as AI),

and False Negatives (images incorrectly classified as non-AI). A detailed categorization of the results is shown in the Annex (Table A1).

For additional visual representations of the AI detection results, including screenshots, please follow the link provided below or scan the QR code using your mobile device:

Link (https://drive.google.com/drive/folders/1E9fhsYvNzYNzMRYHH_W89fCaX2-epND-)

To gain a deeper understanding of the AI tools' performance, Table 1 illustrates a summary of the experimental outcomes. Figure 1 visually summarizes the detection results using bar charts, showing the distributions of true positives, true negatives, false positives, and false negatives for each tool across the 122-image dataset. Additionally, Table 2 delves into the specific performance metrics calculated for each AI tool.

Table 1. Comparison of AI Detection Results by 3 Different Tools

Detection Results	AI or Not	Hugging Face	Hive Moderation
True Positive	55	5	59
True Negative	61	34	61
False Positive	0	27	0
False Negative	6	56	2

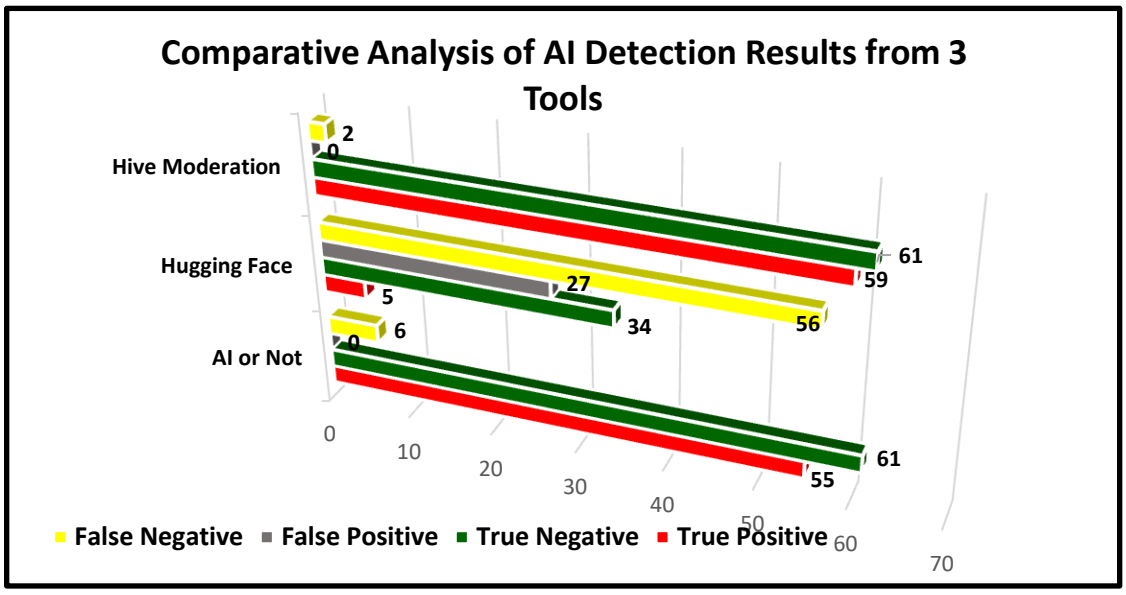

Figure 1. Comparison of AI Detection Results by Three (3) Different Tools.

Table 2. Performance Metrics for AI Detection Tools

Metric / Tool	Ai or not	Hugging Face	Hive Moderation
Accuracy	95%	31.9%	98%
Precision	100%	15.6%	100%
Recall	90%	8%	96.7%
F1 Score	94.7%	10.6%	98%
False Positive Rate	0%	44.%	0%
False Negative Rate	9.8%	91.8%	3.2%

Table 2 compares accuracy, precision, recall, F1-Score, false-positive rate, and false-negative rate for three AI detection tools: AI or Not, Hugging Face, and Hive Moderation.

Upon examining the results, our assessment reveals distinct strengths and weaknesses for each tool:

AI or Not:

Strengths: This tool excels in terms of accuracy, precision, and recall. The F1 score indicates a strong balance between these metrics. It also boasts a perfect false-positive rate, indicating that no non-AI images are incorrectly identified.

Weaknesses: While the false-negative rate is moderate, indicating a few missed AI images, further investigation is needed regarding the single "Undetermined" image classification. Additionally, this tool reported false positives for specific images classified as NSFW, suggesting potential improvements in this area.

Hugging Face:

Weaknesses: This tool performs poorly across most metrics, with low accuracy, precision, and recall. The F1 score indicates a poor balance between precision and recall, with both false-positive and false-negative rates high, suggesting significant issues with prediction quality.

Hive Moderation:

Strengths: Similar to "AI or Not," this tool showcases exceptional performance with high accuracy, precision, and recall. The F1 score demonstrates a strong balance, and the low false-positive and false-negative rates indicate reliable predictions.

4.2 Experiment 2: Analyzing Exported WhatsApp Chat Data for Text Steganography: Observations

This research aimed to explore the potential of AI models to identify text steganography in exported WhatsApp chat data. That involved analyzing the data for unusual characters or patterns that might indicate hidden messages.

While limitations in building and training the AI model prevented further exploration with that approach, significant observations were made during the initial investigation:

4.2.1 Hidden Message Visibility

Copying text directly from WhatsApp and pasting it into the designated application (for decoding steganography messages) revealed hidden information. However, copying from the exported chat data and pasting it into the same application only displayed weird symbols, not the decoded message. That suggests that the chat export process might alter or remove the necessary encoding structure for decoding.

4.2.2 Arabic Text Anomaly

Certain Arabic sentences in the original WhatsApp chat appeared unusual, containing a mix of characters and symbols. This observation highlights the importance of examining text anomalies, especially in non-Latin character sets, when investigating potential steganography.

4.3 Experiment 3: The Role of AI in URL Phishing Detection

This research investigated the potential of using Artificial Intelligence (AI) for detecting phishing URLs. A SWOT analysis was carried out to evaluate the strengths, weaknesses, opportunities, and threats associated with that approach.

4.3.1 Strengths

❖ High Accuracy: AI can analyze vast datasets, leading to accurate identification of phishing URLs and reducing false positives and negatives (Basit et al., 2021).

❖ Speed: AI algorithms enable rapid detection of phishing threats, facilitating real-time prevention (ibid).

❖ Adaptability: AI models can continuously learn and adapt to new phishing techniques, ensuring long-term effectiveness (Alkhalil, Hewage, Nawaf, & Khan, 2021).

4.3.2 Weaknesses

❖ Data Dependence: AI models require large and diverse training datasets, which might not always be readily available (Basit et al., 2021).

❖ False Positives: Improperly tuned AI models can generate false positives, leading to unnecessary alerts and disruptions (Alkhalil et al., 2021).

❖ Resource-Intensity: Implementing AI requires significant computational resources and expertise, potentially limiting its accessibility (Ntalampiras, Misuraca, & Rossel, 2023).

4.3.3 Opportunities

❖ Advanced Threat Detection: AI can identify sophisticated and targeted phishing attacks that might bypass traditional methods (Brundage, Avin, Clark, et al., 2018).

❖ Automation: AI can automate threat response and improve incident handling, saving time and reducing human error (ibid).

❖ Enhanced User Training: AI can personalize user training and awareness programs, fostering better recognition and mitigation of phishing attempts (Ntalampiras et al., 2023).

4.3.4 Threats

❖ Adversarial Attacks: Malicious actors might exploit AI to create more convincing phishing attempts, posing a significant threat (Brundage, Avin, Clark, et al., 2018).

❖ Data Privacy Concerns: Using AI might involve processing sensitive data, raising privacy concerns and potential regulatory issues (Ntalampiras et al., 2023).

❖ Over-dependence on AI: Excessive reliance on AI could create security vulnerabilities and hinder adaptation to unforeseen threats (Brundage et al., 2018).

5. Discussion

This research aimed to investigate three research objectives: Namely, detecting AI-generated images and the potential of AI algorithms to distinguish computer-generated visuals from authentic photographs; uncovering hidden messages and AI tools' ability to uncover hidden communication embedded within seemingly harmless text data; and identifying phishing URLs by discussing AI models' ability to accurately differentiate malicious phishing attempts disguised as legitimate websites. The first two objectives were investigated experimentally. Outcomes have shown that the three chosen AI tools for the first experiment and the tool chosen for the second experiment successfully categorized images according to agreed-upon metrics (Gillham, 2023; 2024) and detected hidden messages. As for the third objective, it was not tested experimentally due to a lack of training data. However, a qualitative SWOT analysis was performed to describe potential benefits and challenges as reported recently by many researchers (Brundage et al., 2018; Basit et al., 2021; Alkhalil et al., 2021; Ntalampiras et al., 2023).

This research explores the potential of AI to revolutionize digital forensics. While the findings highlight its potential benefits, it's crucial to acknowledge the limitations and challenges that need to be addressed for widespread adoption.

5.1 Limitations and Challenges

1. Lack of Expertise: The research faced limitations due to the absence of specialized knowledge in training AI models. This highlights the need for collaboration with AI experts to unlock the full potential of AI in this field.
2. Data Access: The lack of access to security centers limited the ability to collect on-site data. Future research should explore alternative data sources or collaborations with relevant institutions.
3. Ethical Concerns: While acknowledged, the ethical and legal complexities surrounding AI and digital forensics require further exploration and development of appropriate solutions.
4. Standardization and Training Data: The absence of standardized frameworks and limited availability of labeled training data pose additional challenges. Standardized methods and ethical, secure labeling are crucial for robust AI implementation.

5. Interpretability and Complexity: A "black box" nature of certain AI algorithms makes it difficult to understand their decision-making process. Additionally, identifying AI systems responsible for incidents can be challenging, further complicating investigations.

5.2 Future Research Directions

Each experiment conducted within this research suggests specific avenues for future exploration:

- Experiment 1: AI-Generated Images:
  - Training specialized AI models to improve detection accuracy.
  - Investigating emerging challenges and threats related to detecting evolving AI-generated content.
- Experiment 2: AI-Based Steganography Analysis:
  - Developing AI models to identify specific symbols or patterns potentially indicating hidden messages within chat data.
- Experiment 3: AI-Phishing Detection:
  - Creating user-friendly interfaces for AI phishing detection tools to broaden accessibility.
  - Continuously improving the accuracy of AI phishing detection through advancements in AI technology.

5.4 Recommendations for Advancing AI in Digital Forensics

Harnessing the potential of AI while addressing its challenges, the following recommendations are proposed (Jarrett & Choo, 2021):

- Invest in purpose-built AI tools: Develop AI specifically designed for digital forensics tasks, such as analyzing large datasets, identifying AI-generated content, and even predicting cybercrime.
- Mitigate Bias and Ensure Fairness: Implement safeguards to prevent bias in AI algorithms and ensure fair and transparent processes while protecting privacy and data rights.
- Foster Collaboration: Encourage collaboration between AI researchers and digital forensics practitioners to ensure AI solutions meet the needs of real-world investigations.
- Educate Digital Forensics Professionals: Provide comprehensive training programs to equip digital forensics professionals with the necessary skills to integrate AI into their work effectively.
- Embrace Continuous Learning: Stay informed about the latest advancements in AI and related technologies to adapt forensic methodologies and effectively address evolving cyber threats.
- Explore Practical Implementation in Lebanon: Prioritize exploring the practical application of AI in the context of Lebanon's unique digital forensics landscape through collaboration with local security agencies and professionals.

By implementing these recommendations, the digital forensics community in Lebanon can position itself as a leader in the responsible adoption and effective utilization of AI in this critical field.

6. Conclusion

Our research explored how AI is used in digital forensics, including detecting fake images, uncovering hidden messages, and identifying phishing scams. While AI shows great promise, there are hurdles to

overcome. Training AI effectively and getting the right data can be tricky, and unintended mistakes can occur.

Overall, this research highlights the potential of AI to revolutionize digital forensics, while also emphasizing the need for ongoing development and refinement to truly combat evolving cyber threats. The future of digital forensics appears increasingly intertwined with AI, and this research paves the way for further exploration in this exciting field.

References

Abbas, A., & Hurr, R. (2023, December). A Survey of Recent Advances in Artificial Intelligence machine learning and deep learning to natural language processing. *Pakistan Journal of Linguistics, 5*(1), 102-117. https://doi.org/10.13140/RG.2.2.31157.86249

AI or Not. (2024). *Verification with AI Detection*. Retrieved May 31, 2026, from https://www.aiornot.com/

Alaa El-Din, E. A. (2022). Artificial Intelligence in Forensic Science: Invasion or Revolution? *Egypt. Soc. Clin. Toxicol. J., 10* (2), 20–32. https://doi.org/10.21608/esctj.2022.158178.1012

Alkhalil, Z., Hewage, C., Nawaf, L., & Khan, I. (2021). Phishing Attacks: A Recent Comprehensive Study and a New Anatomy. *Frontiers in Computer Science, 3*, 563060. https://doi.org/10.3389/fcomp.2021.563060

Arora, T., & Soni, R. (2021). A Review of Techniques to Detect the GAN-Generated Fake Images. In *Generative Adversarial Networks for Image-to-Image Translation*; Solanki, A., Nayyar, A., Naved, M., Eds., (pp. 125-159). Chapter 6. Academic Press. https://doi.org/10.1016/B978-0-12-823519-5.00004-X

Balogun, A. M., & Zuva, T. (2017). Open Ethical Issues in Digital Forensic Systems. *International Journal of EBusiness and EGovernment Studies*, *9*(1), 55-69. https://izlik.org/JA64SB79CA

Basit, A., Zafar, M., Liu, X., Javed, A. R., Jalil, Z., & Kifayat, K. A. (2021). Comprehensive Survey of AI-Enabled Phishing Attacks Detection Techniques. *Telecomm. Syst., 76*(1), 139–154. https://doi.org/10.1007/s11235-020-00733-2

Bird, J. J., & Lotfi, A. (2024, February). CIFAKE: Image Classification and Explainable Identification of AI-Generated Synthetic Images. *IEEE ACCESS, 12*, 15642-15650. https://doi.org/10.1109/ACCESS.2024.3356122

Bouijij, H., & Berqia, A. (2021). Machine Learning Algorithms Evaluation for Phishing URL Classification. In *Proceedings of the 2021 4th International Symposium on Advanced Electrical and Communication Technologies* (ISAECT), (pp. 01–05). https://doi.org/10.1109/ISAECT53699.2021.9668489

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G., Steinhardt, J., Flynn, C., hÉigeartaigh, S., Beard, S., Belfield, H., Farquhar, S., & Amodei, D. (2018). *The Malicious Use of Artificial Intelligence:*

Forecasting, Prevention, and Mitigation (pp. 1-101). https://doi.org/10.48550/arXiv.1802.07228

Chinnikatti, S. (2018). Artificial Intelligence in Forensic Science. *Forensic Sci. Addict. Res., 2*, 000554 https://doi.org/10.31031/FSAR.2018.03.000554

Costantini, S., De Gasperis, G., & Olivieri, R. (2019). Digital Forensics and Investigations Meet Artificial Intelligence. *Ann. Math. Artif. Intell., 86*(1), 193–229. https://doi.org/10.1007/s10472-019-09632-y

De Rezende, E. R. S., Ruppert, G. C. S., & Carvalho, T. (2017). Detecting Computer Generated Images with Deep Convolutional Neural Networks. *Proceedings of the 30th SIBGRAPI Conference on Graphics, Patterns and Images* (SIBGRAPI), (pp. 71-78). Niteroi, Brazil. https://doi.org/10.1109/SIBGRAPI.2017.16

Elsaesser, C., & Tanner, M. C. (2001). *Automated diagnosis for computer forensics*. Retrieved May 25, 2026, from https://www.mitre.org/sites/default/files/pdf/esaesser_forensics.pdf

Fayyad-Kazan, H., Hejase, H. J., El Dirani, A., & Rkein, H. (2021). JPEG Steganography: Hiding in Plain Sight. *International Journal of Forensic Sciences, 6*(1), 1-11. https://doi.org/10.23880/ijfsc-16000223

Fayyad-Kazan, H., Hejase, H. J., Darwish, C. D., & & Hejase, A. J. (2024, September). A Pilot Study to Assess the Success Rate of Email Scams by Phishing: Case in Lebanon. *Contemporary Studies in Applied Sciences, 1*(1), 1-21. https://doi.org/10.20849/csas.v1i1.1453

Franke, K., Hjelmås, E., & Wolthusen, S.D. (2013). Advancing Digital Forensics. In: Dodge, R.C., Futcher, L. (eds) *Information Assurance and Security Education and Training* (pp. 288-295). WISE 2009, WISE 2011, WISE 2013. IFIP Advances in Information and Communication Technology, vol. 406. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-39377-8_34

Galego Hernandes, P. R., Floret, C. P., Cardozo De Almeida, K. F., Da Silva, V. C., Papa, J. P., & Pontara Da Costa, K. A. (2021). Phishing Detection Using URL-Based XAI Techniques. In *Proceedings of the 2021 IEEE Symposium Series on Computational Intelligence* (SSCI), (pp. 01–06). https://doi.org/10.1109/SSCI50451.2021.9659981

Gillham, J. (2023, August 25). AI Content Detector False Positives – Accused of Using Chat GPT Or Other AI? *Blog*. Retrieved June 5, 2026, from https://originality.ai/blog/ai-content-detector-false-positives

Gilham, J. (2024, March 27). AI Content Detector Accuracy Review + Open Source Dataset and Research Tool *Blog*. Retrieved June 5, 2026, from https://originality.ai/blog/ai-content-detection-accuracy

Göring, S., Ramachandra Rao, R. R., Merten, R., & Raake, A. (2023). Appeal and Quality Assessment for AI-Generated Images. *Proceedings of the 15th International Conference on Quality of Multimedia Experience* - 2023 (QoMEX), Ghent, Belgium 2023, pp 115–118. https://doi.org/10.1109/QoMEX58391.2023.10178486

Hawking, S. (2016, October 20). Stephen Hawking warns of dangerous AI. *BBC*. Retrieved May 25,

2026, from https://www.bbc.com/news/av/technology-37713942

Hive Moderation. (2024). *Machine learning models to detect AI-generated content*. Retrieved May 31, 2026, from https://hivemoderation.com/ai-generated-content-detection

Huggingface. (2024). *Maybe's AI Art Detector*. Retrieved May 31, 2026, from https://huggingface.co/spaces/umm-maybe/AI-image-detector

Hussein, F., & Hejase, H.J. (2022). Artificial Intelligence and Fake News. *Žurnalistikos Tyrimai, 16*, 39–71. https://doi.org/10.15388/ZT/JR.2022.2

Irons, A., & Lallie, H. S. (2014). Digital Forensics to Intelligent Forensics. *Future Internet, 6*(3), 584–596. https://doi.org/10.3390/fi6030584

Jarrett, A., & Choo, K.-K. R. (2021). The Impact of Automation and Artificial Intelligence on Digital Forensics. Wiley Interdiscip. *Rev. Forensic Sci., 3*, 1-17. https://doi.org/10.1002/wfs2.1418

Khan, M., Ali, S., & Asghar, Z. (2015). Introduction to Linguistic Steganography. *Nonlinear Eng., 4*. https://doi.org/10.1515/nleng-2015-0013.

Kominkova Oplatkova, Z., Hološka, J., Zelinka, I., & Senkerik, R. (2008). Steganography Detection using Neural Networks. *Proceedings of the 19th International Workshop on Database and Expert Systems Applications*, (pp. 571-575). Turin, Italy. https://doi.org/10.1109/DEXA.2008.82

Krivchenkov, A., Misnevs, B., & Pavlyuk, D. (2019). Intelligent Methods in Digital Forensics: State of the Art. In *Reliability and Statistics in Transportation and Communication*; Kabashkin, I., Yatskiv (Jackiva), I., Prentkovskis, O., Eds.; Lecture Notes in Networks and Systems, (pp 274–284). Springer International Publishing: Cham. https://doi.org/10.1007/978-3-030-12450-2_26

Lamberti, A. (2023, July 31). Tackling bias in large ML models: the role of synthetic data. *Syntheticus*. Retrieved May 31, 2026, from https://syntheticus.ai/blog/tackling-bias-in-large-ml-models-the-role-of-synthetic-data

Li, S., Wang, J., & Liu, P. (2023). Detection of Generative Linguistic Steganography Based on Explicit and Latent Text Word Relation Mining Using Deep Learning. *IEEE Trans. Dependable Secure Comput, 20*(2), 1476–1487. https://doi.org/10.1109/TDSC.2022.3156972

MacDermott, A., Baker, T., & Shi, Q. (2018). Iot Forensics: Challenges for the Ioa Era. *Proceedings of the 2018 9th IFIP International Conference on New Technologies, Mobility and Security* (NTMS), (pp. 1-5). Paris, France. https://doi.org/10.1109/NTMS.2018.8328748

Maratsi, M. I., Popov, O., Alexopoulos, C., & Charalabidis, Y. (2022). Ethical and Legal Aspects of Digital Forensics Algorithms: The Case of Digital Evidence Acquisition. In *Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance*; ICEGOV ’22 (pp. 32-40). Association for Computing Machinery: New York, NY, USA. https://doi.org/10.1145/3560107.3560114

Matthew Maybe (2022, October 15). Can an AI learn to identify “AI art”? *Blog* Retrieved May 31, 2026, from https://medium.com/@matthewmaybe/can-an-ai-learn-to-identify-ai-art-545d9d6af226

Mitchell, F. (2010). The Use of Artificial Intelligence in Digital Forensics: An Introduction. *Digit. Evid.*

Electron. Signat. Law Rev., 7, 35-41. https://doi.org/10.14296/deeslr.v7i0.1922

Mohammed, H., Clarke, N., & Li, F. (2016). An Automated Approach for Digital Forensic Analysis of Heterogeneous Big Data. *J. Digit. Forensics Secur. Law, 11*(2), 137-152. https://doi.org/10.15394/jdfsl.2016.1384

Muthukrishnan, N., Maleki, F., Ovens, K., Reinhold, C., Forghani, B., Forghani, R. (2020). Brief History of Artificial Intelligence. *Neuroimaging Clin. N. Am., 30* (4), 393–399. https://doi.org/10.1016/j.nic.2020.07.004.

Ntalampiras, S., Misuraca, G., & Rossel, P. (2022). Artificial Intelligence and Cybersecurity Research. *European Union Agency for Cybersecurity* (ENISA). Retrieved June 5, 2026, from https://www.enisa.europa.eu/publications/artificial-intelligence-and-cybersecurity-research

Schneider, J., & Breitinger, F. (2023). Towards AI Forensics: Did the Artificial Intelligence System Do It? *J. Inf. Secur. Appl., 76*, 103517. https://doi.org/10.1016/j.jisa.2023.103517

Silberg, J., & Manyika, J. (2019, June 6). Tackling bias in artificial intelligence (and in humans). *McKinsey Global Institute*. Retrieved May 31, 2026, from https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans

Solanke, A. A. (2022). Explainable digital forensics AI: Towards mitigating distrust in AI-based digital forensics analysis using interpretable models. *Forensic Science International: Digital Investigation, 42*, Supplement, 301403. https://doi.org/10.1016/j.fsidi.2022.301403

Stahl, B., Carroll-mayer, M., Elizondo, D., Wakunuma, K., & Zheng, Y. (2012). Intelligence Techniques in Computer Security and Forensics: At the Boundaries of Ethics and Law. *Comput. Intell. Priv. Secur., 394*, 237–258. https://doi.org/10.1007/978-3-642-25237-2_14

Yang, Z., Wang, K., Li, J., Huang, Y., & Zhang, Y.-J. (2019). TS-RNN: Text Steganalysis Based on Recurrent Neural Networks. *IEEE Signal Process. Lett., 26*(12), 1743–1747. https://doi.org/10.1109/LSP.2019.2920452

Zhang, Z., Hamadi, H. A., Damiani, E., Yeun, C. Y., & Taher, F. (2022). Explainable Artificial Intelligence Applications in Cyber Security: State-of-the-Art in Research. *IEEE Access, 10*, 93104–93139. https://doi.org/10.1109/ACCESS.2022.3204051.

Annex

Table A1. The results of AI detection for selected images using three different AI tools

Image	AI%	AI tool 1: AI or Not	AI tool 2: Hugging Face	AI tool 3: Hive Modification	Detection Result (Tool 1)	Detection Result (Tool 2)	Detection Result (Tool 3)
IMG001	100 %AI	100% AI	45% AI	99.9% AI	True Positive	False Negative	True Positive
IMG002	100 %AI	100% AI	69% AI	99.9% AI	True Positive	False Negative	True Positive
IMG003	100 %AI	100% AI	1% AI	99.9% AI	True Positive	False Negative	True Positive
IMG004	100 %AI	100% AI	28% AI	98.3% AI	True Positive	False Negative	True Positive
IMG005	100 %AI	100% AI	27% AI	99.9% AI	True Positive	False Negative	True Positive
IMG006	100 %AI	100% AI	52% AI	99.9% AI	True Positive	False Negative	True Positive
IMG007	100 %AI	100% AI	36% AI	99.9% AI	True Positive	False Negative	True Positive
IMG008	100 %AI	100% AI	1% AI	99.9% AI	True Positive	False Negative	True Positive
IMG009	100 %AI	100% AI	20% AI	98.3% AI	True Positive	False Negative	True Positive
IMG010	100 %AI	100% AI	12% AI	99.9% AI	True Positive	False Negative	True Positive
IMG011	100 %AI	100% AI	36% AI	99.9% AI	True Positive	False Negative	True Positive
IMG012	100 %AI	100% AI	3% AI	99.9% AI	True Positive	False Negative	True Positive
IMG013	100 %AI	100% AI	18% AI	99.9% AI	True Positive	False Negative	True Positive
IMG014	100 %AI	100% AI	6% AI	99.9% AI	True Positive	False Negative	True Positive
IMG015	100 %AI	100% AI	14% AI	99.9% AI	True Positive	False Negative	True Positive
IMG016	100 %AI	100% AI	60% AI	99.9% AI	True Positive	False Negative	True Positive
IMG017	100 %AI	100% AI	87% AI	99.9% AI	True Positive	False Negative	True Positive
IMG018	100 %AI	100% AI	96% AI	99.9% AI	True Positive	True Positive	True Positive
IMG019	100 %AI	100% AI	96% AI	99.9% AI	True Positive	True Positive	True Positive
IMG020	100 %AI	100% AI	94% AI	99.9% AI	True Positive	True Positive	True Positive
IMG021	100 %AI	100% AI	5% AI	99.9% AI	True Positive	False Negative	True Positive
IMG022	100 %AI	100% AI	7% AI	99.6% AI	True Positive	False Negative	True Positive
IMG023	100 %AI	0% AI	48% AI	79% AI	False Negative	False Negative	False Negative
IMG024	100 %AI	0% AI	22% AI	99.9% AI	False Negative	False Negative	True Positive
IMG025	100 %AI	0% AI	16% AI	99.9% AI	False Negative	False Negative	True Positive
IMG026	100 %AI	0% AI	19% AI	99.9% AI	False Negative	False Negative	True Positive
IMG027	100 %AI	100% AI	56% AI	99.6% AI	True Positive	False Negative	True Positive
IMG028	100 %AI	100% AI	89% AI	99.9% AI	True Positive	False Negative	True Positive
IMG029	100 %AI	100% AI	89% AI	99.9% AI	True Positive	False Negative	True Positive
IMG030	100 %AI	Not identified	41% AI	99.9% AI	True Positive	False Negative	True Positive

IMG031	100 %AI	100% AI	24% AI	99.9% AI	True Positive	False Negative	True Positive
IMG032	100 %AI	100% AI	2% AI	99.9% AI	True Positive	False Negative	True Positive
IMG033	100 %AI	100% AI	3% AI	99.9% AI	True Positive	False Negative	True Positive
IMG034	100 %AI	100% AI	3% AI	99.9% AI	True Positive	False Negative	True Positive
IMG035	100 %AI	100% AI	5% AI	99.9% AI	True Positive	False Negative	True Positive
IMG036	100 %AI	100% AI	45% AI	99.9% AI	True Positive	False Negative	True Positive
IMG037	100 %AI	100% AI	89% AI	99.9% AI	True Positive	False Negative	True Positive
IMG038	100 %AI	100% AI	93% AI	99.9% AI	True Positive	True Positive	True Positive
IMG039	100 %AI	100% AI	78% AI	99.8% AI	True Positive	False Negative	True Positive
IMG040	100 %AI	0% AI	21% AI	60.5% AI	False Negative	False Negative	False Negative
IMG041	100 %AI	100% AI	3% AI	99.9% AI	True Positive	False Negative	True Positive
IMG042	100 %AI	0% AI	2% AI	96.8% AI	False Negative	False Negative	True Positive
IMG043	100 %AI	100% AI	11% AI	99.9% AI	True Positive	False Negative	True Positive
IMG044	100 %AI	100% AI	37% AI	99% AI	True Positive	False Negative	True Positive
IMG045	100 %AI	100% AI	24% AI	99.9% AI	True Positive	False Negative	True Positive
IMG046	100 %AI	100% AI	33% AI	99.9% AI	True Positive	False Negative	True Positive
IMG047	100 %AI	100% AI	8% AI	99.9% AI	True Positive	False Negative	True Positive
IMG048	100 %AI	100% AI	20% AI	99.9% AI	True Positive	False Negative	True Positive
IMG049	100 %AI	100% AI	18% AI	97.7% AI	True Positive	False Negative	True Positive
IMG050	100 %AI	100% AI	58% AI	99.9% AI	True Positive	False Negative	True Positive
IMG051	100 %AI	100% AI	88% AI	99.9% AI	True Positive	False Negative	True Positive
IMG052	100 %AI	100% AI	81% AI	99.9% AI	True Positive	False Negative	True Positive
IMG053	100 %AI	100% AI	6% AI	99.9% AI	True Positive	False Negative	True Positive
IMG054	100 %AI	100% AI	27% AI	99.9% AI	True Positive	False Negative	True Positive
IMG055	100 %AI	100% AI	52% AI	99.9% AI	True Positive	False Negative	True Positive
IMG056	100 %AI	100% AI	8% AI	99.6% AI	True Positive	False Negative	True Positive
IMG057	100 %AI	100% AI	95% AI	99.2% AI	True Positive	True Positive	True Positive
IMG058	100 %AI	100% AI	6% AI	99.9% AI	True Positive	False Negative	True Positive
IMG059	100 %AI	100% AI	3% AI	99.9% AI	True Positive	False Negative	True Positive
IMG060	100 %AI	100% AI	82% AI	99.9% AI	True Positive	False Negative	True Positive
IMG061	100 %AI	100% AI	52% AI	99.9% AI	True Positive	False Negative	True Positive
IMG062	0%AI	0% AI	32% AI	1.3% AI	True Negative	False Positive	True Negative
IMG063	0%AI	0% AI	8% AI	0% AI	True Negative	True Negative	True Negative
IMG064	0%AI	0% AI	19% AI	0% AI	True Negative	False Positive	True Negative
IMG065	0%AI	0% AI	18% AI	0.3% AI	True Negative	False Positive	True Negative

IMG066	0%AI	0% AI	58% AI	2.4% AI	True Negative	False Positive	True Negative
IMG067	0%AI	0% AI	3% AI	0.3% AI	True Negative	True Negative	True Negative
IMG068	0%AI	0% AI	3% AI	0.4% AI	True Negative	True Negative	True Negative
IMG069	0%AI	0% AI	11% AI	2.7% AI	True Negative	False Positive	True Negative
IMG070	0%AI	0% AI	82% AI	0% AI	True Negative	False Positive	True Negative
IMG071	0%AI	0% AI	34% AI	0.2% AI	True Negative	False Positive	True Negative
IMG072	0%AI	0% AI	18% AI	0.6% AI	True Negative	False Positive	True Negative
IMG073	0%AI	0%AI	23% AI	0% AI	True Negative	False Positive	True Negative
IMG074	0%AI	0%AI	5% AI	0% AI	True Negative	True Negative	True Negative
IMG075	0%AI	0%AI	83% AI	0% AI	True Negative	False Positive	True Negative
IMG076	0%AI	0%AI	15% AI	1.6% AI	True Negative	False Positive	True Negative
IMG077	0%AI	0%AI	5% AI	0% AI	True Negative	True Negative	True Negative
IMG078	0%AI	0%AI	8% AI	0% AI	True Negative	True Negative	True Negative
IMG079	0%AI	0%AI	16% AI	0% AI	True Negative	False Positive	True Negative
IMG080	0%AI	0%AI	5% AI	0.2% AI	True Negative	True Negative	True Negative
IMG081	0%AI	0%AI	9% AI	0% AI	True Negative	True Negative	True Negative
IMG082	0%AI	0%AI	23% AI	0% AI	True Negative	False Positive	True Negative
IMG083	0%AI	0%AI	2% AI	1.3% AI	True Negative	True Negative	True Negative
IMG084	0%AI	0%AI	8% AI	0.5% AI	True Negative	True Negative	True Negative
IMG085	0%AI	0%AI	11% AI	0% AI	True Negative	False Positive	True Negative
IMG086	0%AI	0%AI	1% AI	0% AI	True Negative	True Negative	True Negative
IMG087	0%AI	0%AI	1% AI	0% AI	True Negative	True Negative	True Negative
IMG088	0%AI	0%AI	17% AI	0% AI	True Negative	False Positive	True Negative
IMG089	0%AI	0%AI	1% AI	0% AI	True Negative	True Negative	True Negative
IMG090	0%AI	0%AI	6% AI	0% AI	True Negative	True Negative	True Negative
IMG091	0%AI	0%AI	0% AI	0% AI	True Negative	True Negative	True Negative
IMG092	0%AI	0%AI	1% AI	0% AI	True Negative	True Negative	True Negative
IMG093	0%AI	0%AI	7% AI	0.3% AI	True Negative	True Negative	True Negative
IMG094	0%AI	0%AI	3% AI	0.1% AI	True Negative	True Negative	True Negative
IMG095	0%AI	0%AI	6% AI	0% AI	True Negative	True Negative	True Negative
IMG096	0%AI	0%AI	24% AI	0.1% AI	True Negative	False Positive	True Negative
IMG097	0%AI	0%AI	7% AI	0.2% AI	True Negative	True Negative	True Negative
IMG098	0%AI	0%AI	3% AI	0% AI	True Negative	True Negative	True Negative
IMG099	0%AI	0%AI	45% AI	0.1% AI	True Negative	False Positive	True Negative
IMG100	0%AI	0%AI	30% AI	0% AI	True Negative	False Positive	True Negative

IMG101	0%AI	0% AI	0% AI	0.1% AI	True Negative	True Negative	True Negative
IMG102	0%AI	0% AI	5% AI	0% AI	True Negative	True Negative	True Negative
IMG103	0%AI	0% AI	1% AI	0.1% AI	True Negative	True Negative	True Negative
IMG104	0%AI	0% AI	0% AI	0% AI	True Negative	True Negative	True Negative
IMG105	0%AI	0% AI	3% AI	0% AI	True Negative	True Negative	True Negative
IMG106	0%AI	0% AI	0% AI	0% AI	True Negative	True Negative	True Negative
IMG107	0%AI	0% AI	1% AI	0% AI	True Negative	True Negative	True Negative
IMG108	0%AI	0% AI	7% AI	0% AI	True Negative	True Negative	True Negative
IMG109	0%AI	0% AI	3% AI	0.2% AI	True Negative	True Negative	True Negative
IMG110	0%AI	0% AI	2% AI	0% AI	True Negative	True Negative	True Negative
IMG111	0%AI	0% AI	3% AI	0.1% AI	True Negative	True Negative	True Negative
IMG112	0%AI	0% AI	38% AI	0% AI	True Negative	False Positive	True Negative
IMG113	0%AI	0% AI	11% AI	0.1% AI	True Negative	False Positive	True Negative
IMG114	0%AI	0% AI	6% AI	1.5% AI	True Negative	True Negative	True Negative
IMG115	0%AI	0% AI	22% AI	0% AI	True Negative	False Positive	True Negative
IMG116	0%AI	0% AI	46% AI	0% AI	True Negative	False Positive	True Negative
IMG117	0%AI	0% AI	35% AI	0% AI	True Negative	False Positive	True Negative
IMG118	0%AI	0% AI	36% AI	0% AI	True Negative	False Positive	True Negative
IMG119	0%AI	0% AI	45% AI	0% AI	True Negative	False Positive	True Negative
IMG120	0%AI	0% AI	4% AI	0.2% AI	True Negative	True Negative	True Negative
IMG121	0%AI	0% AI	15% AI	0% AI	True Negative	False Positive	True Negative
IMG122	0%AI	0% AI	90% AI	0% AI	True Negative	False Positive	True Negative

Table A1 presents the results of AI detection for selected images using three different AI tools: "AI or not," "hugging face," and "hive modification." The table includes information on the estimated percentage of AI in each image as analyzed by the three tools and the detection results, including true positives, true negatives, false positives, false negatives, and unclassified.