The reality of the current moment is characterized by the acceleration of the evolution of cyber security, in particular due to the influence of artificial intelligence technologies [1], an increase in the speed of computing processes, a rapid increase in the complexity and volume of data that must be processed and taken into account in the processes of analyzing the state of cyber security, the transition to a risk-oriented approach to cyber security, the correlation of domestic approaches with the international practices of CISA [2], NIST [3], MITER ATT&CK[4], etc.

Each of the participants in public administration should be familiar with the processes of organizing the full cycle of response to cyber incidents, its main components, properties, and resources, and understand their place, functions, and acquire the necessary competencies. The complete cyber incident response cycle consists of several interrelated modules, namely: the preparation, the detection and analysis, the deterrence, the elimination, the recovery, and the analysis of the effectiveness of cyber incident response measures. In fact, it is a cyclical process, the main one of which is the preparation stage, in terms of the amount of resources, time, and qualifications required. The stage is performed almost continuously. If immediate cyber incident response actions begin at the time a cyber incident is initiated, the outcome for the institution will be known to be negative.

1. Hilpisch Y., Ali M.G., Jasim A.K., Abdulrahman S.A.R., Abu-AlShaeer M.J., Almansoori K.W.N., Tregubenko I. Strategic Technological Integration and National Industrial Resilience: Assessing AI-Driven Efficiency Across Critical Sectors. (2026), 2855 CCIS, pp. 507 - 522, DOI: 10.1007/978-3-032-17023-1_30
2. Federal Government Cybersecurity Incident and Vulnerability Response Playbooks. URL: https://www.cisa.gov/resources-tools/resources/federal-government-cybersecurity-incident-and-vulnerability-response-playbooks (application date 07.05.2026).
3. NIST SP 800-53 Security and Privacy Controls for Information Systems and Organizations. URL: https://csrc.nist.gov/pubs/sp/800/53/r5/upd1/final (application date 07.05.2026).
4. Adversarial Tactics, Techniques, and Common Knowledge (MITRE ATT&CK). URL: https://attack.mitre.org (application date 07.05.2026).

Розробка алгоритму виявлення ШІ-згенерованих зображень на основі машинного навчання

УДК 621.395.7 (043.2)

Фляк Владислав

Національний університет "Одеська політехніка",10328099@stud.op.edu.ua

Сьогодні цифрові технології радикально змінили спосіб сприйняття інформації. Фотографії та відео вже не є надійним доказом реальності події, оскільки штучний інтелект здатний створювати повністю синтетичний контент, який важко відрізнити від справжнього. ШІ-згенеровані зображення стали масовим явищем у соціальних мережах, рекламі, медіа та навіть судовій практиці. Це призводить до кризи довіри до візуальної інформації та створює

серйозні загрози для кібербезпеки, журналістики та суспільства загалом. Актуальність теми обумовлена необхідністю розробки надійних алгоритмів автоматичного виявлення такого контенту на основі машинного навчання.

ШІ-згенеровані зображення – це синтетичні візуальні матеріали, створені алгоритмами штучного інтелекту, переважно методами глибокого навчання. На відміну від традиційних фотографій, які фіксують реальний світловий потік сенсором камери, вони формуються математично – шляхом моделювання розподілу пікселів. Сучасні моделі (GAN, Diffusion Models) здатні відтворювати складні структури: людські обличчя з реалістичною мімікою, текстури шкіри, відблиски, природні ландшафти та архітектурні сцени. Рівень деталізації настільки високий, що для неозброєного ока відмінність між реальним і синтетичним зображенням часто є непомітною.

Однак принцип формування таких зображень суттєво відрізняється від фізичної фотографії. Реальні знімки містять унікальний шум сенсора (PRNU), метадані EXIF та фізично обґрунтовані тіні й відблиски. ШІ-зображення таких природних слідів не мають або імітують їх лише приблизно. Вони генеруються з випадкового шуму або латентного вектора, тому завжди несуть статистичні «відбитки» моделі. Згідно з оглядом Verdoliva (2020), навіть найсучасніші генератори залишають артефакти: асиметрію очей, нереалістичні зуби, неправильні переходи текстур або спектральні аномалії в частотному домені.

Основними методами генерації сьогодні є Generative Adversarial Networks (GAN) та Diffusion Models. GAN працюють за принципом змагання двох мереж: генератор створює зображення, а дискримінатор намагається його розпізнати. Diffusion Models діють інакше: спочатку додають шум до зображення, а потім навчаються його прибирати крок за кроком. Саме вони лежать в основі Stable Diffusion, DALL-E та Midjourney і сьогодні дають найкращу якість та гнучкість. Умовна генерація дозволяє створювати зображення за текстовим описом, що робить контент ще більш різноманітним і небезпечним для виявлення.

Існуючі підходи до виявлення можна розділити на пасивні (форензичні) та активні (на основі машинного навчання). Пасивні методи (PRNU, ELA, частотний аналіз) шукають природні сліди камери, але сучасні генератори їх добре імітують. Активні методи використовують згорткові мережі (ResNet, EfficientNet), проте аналізують лише просторові ознаки і не враховують частотні артефакти. Epstein et al. (2023) та Cozzolino et al. (2024) підтверджують, що універсального рішення поки немає, а головною проблемою є слабка генералізація на нові моделі генерації.

Для подолання зазначених обмежень у даній роботі запропоновано гібридний алгоритм виявлення ШІ-згенерованих зображень, що поєднує аналіз просторових та частотних ознак з механізмом уваги. Алгоритм базується на двох паралельних гілках обробки: просторова гілка використовує попередньо натреновану мережу EfficientNet-B0 для витягнення візуальних ознак (краї, текстури, форми), а частотна гілка обчислює дискретне косинусне перетворення (DCT) та витягає спектральні артефакти через окрему згорткову підмережу. Об'єднання ознак здійснюється через механізм Attention Fusion, який динамічно зважує внесок кожної гілки залежно від вхідного зображення. Це є ключовою перевагою над існуючими методами: замість фіксованого об'єднання модель

адаптивно визначає, чи просторові, чи частотні ознаки є більш інформативними для конкретного зразка. Навчання реалізовано зі стратегією transfer learning (заморожування backbone з наступним fine-tuning), LR warmup, cosine annealing та early stopping. Експериментальна оцінка на датасеті GenImage (зображення від 10+ генераторів: Stable Diffusion, GLIDE, BigGAN, StyleGAN3, VQ-Diffusion та ін.) показала accuracy 94.3%, F1-score 0.929, precision 0.907, recall 0.951 та ROC-AUC 0.976, що підтверджує високу ефективність гібридного підходу та його здатність надійно виявляти зображення від різноманітних генеративних моделей.

1. Verdoliva L. Media Forensics and DeepFakes: an overview. arXiv, 2020. URL: https://arxiv.org/pdf/2001.06564
2. Rafique R., Gantassi R., Amin R. та ін. Deep fake detection and classification using error-level analysis and deep learning. Scientific Reports (Nature), 2023. URL: https://www.nature.com/articles/s41598-023-34629-3
3. Epstein D.C., Jain I., Wang O., Zhang R. Online Detection of AI-Generated Images. ICCV Workshop, 2023. URL:https://openaccess.thecvf.com/content/ICCV2023W/DFAD/papers/Epstein_Online_Detection_of_AI-Generated_Images__ICCVW_2023_paper.pdf
4. Cozzolino D., Poggi G., Corvi R., Nießner M., Verdoliva L. Raising the Bar of AI-generated Image Detection with CLIP. CVPR Workshop, 2024. URL:https://openaccess.thecvf.com/content/CVPR2024W/WMF/papers/Cozzolino_Raising_the_Bar_of_AI-generated_Image_Detection_with_CLIP_CVPRW_2024_paper.pdf

Modern data hiding techniques: adaptivity, artificial intelligence and content synthesis

UDC 004.056:004.8 (043.2)

Artem Frolov[1], Vasyl Rizak[2]

Uzhhorod National University,
[1]artem.frolov@uzhnu.edu.ua, [2]vrizak@uzhnu.edu.ua

Modern steganography has evolved from simple data hiding in least significant bits (LSB) to complex methods that exploit machine learning and adaptive algorithms. The aim of this work is to analyze contemporary data hiding techniques that combine adaptive algorithms, neural network architectures and generative models, and to outline promising directions for further research.

1. Adaptive embedding based on distortion minimization. This is the “gold standard” of modern steganography: instead of embedding data uniformly, the algorithm analyses the content and identifies regions where modifications will be least detectable by steganalyzers. The mechanism relies on additive cost functions: each pixel is assigned a modification cost — pixels on object edges and in textured areas have low cost, while smooth surfaces (e.g. clear sky) have high cost. The key algorithms are: 1) S-UNIWARD, which operates in the spatial and frequency