1. Nestulya S., Shara S. Distance learning as a relevant educational technology in higher education institutions, *Scientific Bulletin of Mukachevo State University. Series: Pedagogy and Psychology*, vol. 9, no. 1, pp. 39–46, 2023. doi: 10.52534/msu-pp1.2023.39.
2. Vajjala S., Meurers D. Readability Assessment for Text Simplification: From Analyzing Documents to Identifying Sentential Simplifications, *International Journal of Applied Linguistics*, vol. 165, pp. 159–189, 2014. doi: 10.1075/itl.165.2.04vaj.
3. Zupanc K., Bosnić Z. Automated essay evaluation with semantic analysis, *Knowledge-Based Systems*, vol. 120, pp. 118–132, 2017. doi: 10.1016/j.knosys.2017.01.006.

Алгоритмічні ІПСО: генеративні моделі, мікротаргетинг і захист критичних аудиторій

УДК 004.8

Верголяс О.О.

старший лейтенант, кандидат юридичних наук, старший викладач,
Харківський національний університет повітряних сил ім. І. Кожедуба,
arnijazov@gmail.com, ORCID: https://orcid.org/0000-0002-9780-1298

Розглянуто алгоритмічні ІПСО, що використовують генеративні моделі, синтетичні медіа, bot-swarming і мікротаргетинг. Визначено типові сценарії таких атак, індикатори їх виявлення та базову модель реагування. Обґрунтовано, що протидія ШІ-підсиленим ІПСО потребує не окремих спростувань, а узгодженого циклу моніторингу, правової оцінки, стратегічних комунікацій і міжвідомчого обміну індикаторами [1-5].

Генеративний ШІ змінив не лише інструменти ІПСО, а й темп їх виконання. Якщо раніше інформаційна операція потребувала автора, редактора, монтажера і мережі поширення, то нині значну частину циклу можна автоматизувати: мовна модель готує варіанти тексту, генератор зображень створює візуальний супровід, синтез голосу дає псевдоавтентичне аудіо, а бот- або рекламна інфраструктура перевіряє, який меседж краще діє на конкретну групу. Для України це не абстрактна технологічна тема: російська агресія поєднує кібероперації, дезінформацію, психологічний тиск на родини військовослужбовців і спроби розколоти довіру між військом, владою та суспільством.

Алгоритмічні ІПСО доцільно розглядати як зміну архітектури впливу. Її основними компонентами є synthetic identity, deepfake-контент, LLM-мікротексти та bot-swarming. Синтетичні акаунти імітують локальну присутність і створюють враження органічної дискусії. Deepfake може бути не лише повністю синтетичним відео, а й змішаною формою: справжній фрагмент виступу, змінена аудіодоріжка, обрізаний контекст або поширення через канал із готовою аудиторією [3; 5]. LLM-мікротексти діють тонше: короткі повідомлення для військових, родин військовослужбовців, медиків, енергетиків, волонтерів чи місцевих громад маскуються під коментар, «інсайд» або побутову пораду. Крос-платформні хвилі переносять наратив між Telegram,

короткими відео, коментарями під новинами та псевдомедійними обговореннями.

Виявлення таких операцій не можна будувати на одному «детекторі ШІ». Надійнішою є сукупність сигналів: стилометрія тексту, синхронність публікацій, графові ознаки неорганічного поширення, мультимодальні артефакти і темп зміни наративу. Якщо після спростування повідомлення не зникає, а швидко змінює формулювання, канал або аудиторію, ймовірно, діє адаптивний контур впливу. Такі індикатори мають накопичуватися у hash-банках медіа та бібліотеках ознак, сумісних із практиками NIST і CISA [1-3].

Протидія починається не з публічного спростування, а з процесу. Якщо установа не знає, хто фіксує інцидент, хто проводить первинну перевірку, хто готує повідомлення і хто має право говорити публічно, вона програє перші години. Технічний блок має охоплювати перевірку походження медіа, provenance/watermark-маркування офіційного контенту, пасивне детектування дипфейків і крос-платформну кореляцію. Організаційний блок має спиратися на SOC-StratCom-OSINT-ланцюжок: аналітики фіксують сигнал, OSINT перевіряє маршрут поширення, StratCom готує реакцію, юридичний блок оцінює ризики обмеження контенту або звернення до платформи [2-4].

Для України раціональною є тришарова модель реагування. Тактичний рівень (0-2 години) забезпечує моніторинг, первинну верифікацію і фіксацію цифрових артефактів. Оперативний рівень (2-24 години) відповідає за кореляцію індикаторів, запуск контрнаративів, юридичну оцінку і координацію із ЗСУ, СБУ, НКЦК та іншими органами. Стратегічний рівень охоплює міжвідомчий обмін індикаторами, зв'язок із платформами, міжнародними партнерами та аналіз тенденцій [2; 4]. Для військових, родин військових, медиків, енергетиків, освітян і місцевих громад потрібні різні канали довіри; тому trusted voices має бути мережею локально авторитетних комунікаторів, а pre-bunking - способом пояснити типові маніпуляції до того, як конкретний фейк набере масштаб.

Практичні кейси підтверджують цю логіку. Deepfake-відео із закликом до «складання зброї» у березні 2022 року було технічно слабким, але одночасне поширення через сайт, телевізійний рядок, Telegram і соціальні мережі створило коротке вікно паніки. DFRLab також описує фальшиві відео із Валерієм Залужним і відеодзвінок із підробленим образом Петра Порошенка для учасників Інтернаціонального легіону; ці випадки важливі переходом від масової дезінформації до вузького таргетування [5].

Отже, ШІ в ІПСО є новою якістю загрози, а не просто новим інструментом для старої пропаганди. Ефективна модель протидії має поєднувати NIST-сумісні технічні практики, FIMI-орієнтовану аналітику, правову пропорційність і локалізовану комунікацію з критичними аудиторіями. Її результатом має бути не ідеальна цензура, а швидке відокремлення небезпечної операції від інформаційного шуму, збереження доказів і точне спростування [1-5].

1. National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). 2023. URL: https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf (дата звернення:

10.11.2025).
2. National Institute of Standards and Technology. Generative AI Profile for the NIST AI RMF (NIST AI 600-1). 2024. URL: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.600-1.pdf (дата звернення: 10.11.2025).
3. NSA; CISA; FBI. Contextualizing Deepfake Threats to Organizations. 2023. URL: https://media.defense.gov/2023/Sep/12/2003298925/-1/-1/0/CSI-deepfakethreats.pdf (дата звернення: 10.11.2025).
4. European External Action Service. 2nd Report on Foreign Information Manipulation and Interference (FIMI). 2024. URL: https://www.eeas.europa.eu/sites/default/files/documents/2024/EEAS-2nd-Report%20on%20FIMI%20Threats-January-2024_0.pdf (дата звернення: 10.11.2025).
5. Osadchuk R. AI tools usage for disinformation in the war in Ukraine. Digital Forensic Research Lab. 09.07.2024. URL: https://dfrlab.org/2024/07/09/ai-tools-usage-for-disinformation-in-the-war-in-ukraine/ (дата звернення: 10.11.2025).

Захист CI/CD-конвеєрів автоматизованого оновлення Docker-контейнерів на основі криптографічного підписування образів

УДК 004.056:004.75

Віталій Тимощук[1], Дмитро Тимощук[2]

Тернопільський національний технічний університет імені Івана Пулюя,
[1]tymoshchuk@tntu.edu.ua, [2]dmytro.tymoshchuk@gmail.com

Розробка програмного забезпечення дедалі частіше базується на використанні контейнеризації та автоматизованих CI/CD-конвеєрів. Docker-контейнери дають змогу швидко збирати, доставляти й оновлювати застосунки, однак водночас створюють нові ризики для безпеки. Особливо критичною є проблема довіри до контейнерного образу, який автоматично завантажується з реєстру та запускається у продуктивному середовищі. У типовій схемі CI/CD новий образ збирається після зміни коду, публікується в container registry, після чого сервер отримує його за тегом, наприклад latest. Такий підхід є зручним, але не гарантує, що завантажений образ справді відповідає легітимній версії застосунку.

Основна загроза полягає в тому, що контейнерний образ може бути підмінений у реєстрі, у процесі доставки або через помилкову конфігурацію CI/CD-конвеєра. У такому випадку продуктивне середовище може запустити шкідливий або застарілий образ як легітимне оновлення. Це безпосередньо пов’язано з атаками на ланцюг постачання програмного забезпечення, кількість яких суттєво зростає внаслідок активного використання автоматизованих засобів збірки, сторонніх залежностей і container registry [1]. Традиційні засоби мережевого захисту не завжди здатні виявити такі атаки, оскільки скомпрометований образ може виглядати як звичайне оновлення.

Метою роботи є підвищення безпеки автоматизованого оновлення Docker-контейнерів шляхом використання криптографічного підписування