https://doi.org/10.31649/1997-9266-2024-172-1-74-80

УДК 004.054

Г. В. Микитин[1]
Х. С. Руда[1]
Ю. Є. Яремчук[2]

МЕТОДОЛОГІЯ БЕЗПЕКИ НЕЙРОМЕРЕЖЕВИХ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ ВИЯВЛЕННЯ DEEPFAKE-МОДИФІКАЦІЙ БІОМЕТРИЧНОГО ЗОБРАЖЕННЯ

[1]Національний університет «Львівська політехніка»;
[2]Вінницький національний технічний університет

Одним з функціональних векторів Стратегії кібербезпеки України є розроблення і впровадження систем захисту різних інформаційних платформ інфраструктури суспільства, зокрема створення безпечних технологій виявлення deepfake-модифікацій біометричного зображення на основі нейронних мереж у кіберпросторі. В роботі досліджено засади безпеки нейромережевих інформаційних технологій (ІТ) у просторі, проблеми deepfake-модифікацій за використання базового підходу до безпечного виявлення deepfake-модифікацій біометричного зображення та методології безпеки багаторівневої нейромережевої ІТ «ресурси – системи – процеси – мережі – управління» згідно з концепцією «об'єкт – загроза – захист». Базовий підхід об'єднує інформаційну нейромережеву технологію та інформаційну технологію підтримки прийняття рішень, структурованими за модульною архітектурою нейромережевої системи виявлення deepfake-модифікацій в просторі «попереднє оброблення даних опрацювання ознак — навчання класифікатора». Ядром методології безпеки ІТ є цілісність функціонування нейромережевої системи виявлення deepfake-модифікацій біометричного зображення обличчя людини і системи аналізу даних, що реалізують інформаційний процес «розділення відеофайлу на кадри — детекція, опрацювання ознак — оцінювання точності класифікатора зображень». Методологія безпеки багаторівневої нейромережевої ІТ ґрунтується на системному та синергетичному підходах, що уможливлюють побудову комплексної системи безпеки ІТ з урахуванням властивості емерджентності за умови впливу ймовірних цілеспрямованих загроз і застосування новітніх технологій протидії на апаратному і програмному рівнях. Запропонована комплексна система безпеки інформаційного процесу виявлення deepfake-модифікацій біометричного зображення охоплює апаратно-програмні засоби відповідно до сегментів: автоматизованої оцінки точності класифікатора; виявлення deepfake-модифікацій в режимі реального часу; послідовного оброблення зображень; оцінювання точності класифікації з використанням хмарних обчислень.

Ключові слова: інтелектуалізація, кібербезпека, біометричне зображення, deepfake, інформаційна технологія, нейронні мережі, система виявлення, базовий підхід, методологія безпеки, комплексна система безпеки.

Вступ

Постановка проблеми. Безпека об'єктів інфраструктури держави у фізичному та кіберпросторі сьогодні є актуальною проблемою у просторі задач інтелектуалізації предметних сфер суспільства. Одним з основних інструментаріїв для вирішення проблеми безпечної інтелектуалізації об'єктів інфраструктури суспільства у просторі задач Індустрії 4.0, Стратегії кібербезпеки України та Європейського акту про стійкість до кіберзагроз (Cyber Resilience Act) є нейромережеві інформаційні технології виявлення deepfake-модифікацій біометричного зображення обличчя людини [1]—[3]. Критерієм точності класифікації біометричних зображень засобами нейронних мереж є безпечність технологій виявлення deepfake-модифікацій, яка обумовлюється методологією безпеки багаторівневої нейромережевої інформаційної технології.

Аналіз останніх досягнень і публікацій. Актуальним є розвиток методологічних засад створення систем кібербезпеки інформаційних технологій функціонування об'єктів інфраструктури суспільства [4], [5]. Сьогодні розгортаються процеси безпеки технологій в задачах виявлення deepfake-модифікацій біометричного зображення облич людини на основі нейронних мереж. Досліджується питання безпеки машинного навчання у сегменті комплексних моделей загроз та відповідних засобів захисту [6], [7]. В роботі [8] розглянуто модель безпеки і конфіденційність даних в глибокому навчанні, як частини машинного навчання у разі впливу відповідних атак. На модель безпеки впливають атаки отруєння в процесі глибокого навчання та атаки ухилення під час прийняття рішення про результат глибокого навчання. На протидію цим атакам передбачені: методи розпізнавання та видалення зловмисних даних; навчання моделі нечутливості до таких даних; маскування структури та параметрів моделі. Конфіденційності даних в процесі глибокого навчання загрожують своєрідні атаки, зокрема інверсія моделі безпеки. Ефективним інструментарієм протидії загрозам конфіденційності є методи криптографії, серед яких гомоморфне шифрування. Цікавим є дослідження питання безпеки апаратного забезпечення глибоких нейронних мереж в просторі «загроза–захист», про що йдеться в праці [9]. Відомі сучасні методи виявлення deepfake-модифікацій біометричного зображення обличчя людини з точністю 0,94…0,99 [10].

Метою роботи є розроблення методології безпеки нейромережевої інформаційної технології згідно з: 1) базовим підходом до виявлення deepfake-модифікацій біометричних зображень; 2) багаторівневою моделлю нейромережевої ІТ: інформаційні ресурси (ІР) – інформаційні системи (ІС) – інформаційні процеси (ІП) – інформаційні мережі (ІМ) – управління інформаційною безпекою (У), яка реалізує конструктивний алгоритм безпечного функціонування нейромережевої ІТ.

Базовий підхід до виявлення deepfake-модифікацій біометричного зображення засобами нейронних мереж

Підґрунтям методології безпеки нейромережевих технологій є базовий підхід до виявлення deepfake-модифікацій біометричного зображення (рис. 1): інформаційна нейромережева технологія (ІТ1); інформаційна технологія підтримки прийняття рішень (ІТ2). Розроблення інформаційних технологій виявлення deepfake-модифікацій біометричного зображення ґрунтується на: використанні підходу поетапного виявлення модифікованих біометричних зображень засобами згорткових нейронних мереж [11]; застосуванні нейромережевої системи виявлення deepfake-модифікацій за її архітектурою та системи підтримки прийняття рішення про якість роботи класифікатора біометричних зображень згідно з методикою оцінювання. В основі інформаційної нейромережевої технології: модель об'єкта, методологія виявлення deepfake-модифікацій, точність класифікації біометричних зображень, методика оцінювання якості роботи класифікатора біометричних зображень. Конструктивний алгоритм ІТ1 «розділення відео на кадри – детекція – опрацювання ознак – класифікація» реалізується архітектурою нейромережевої системи з використанням модульного підходу, що включає

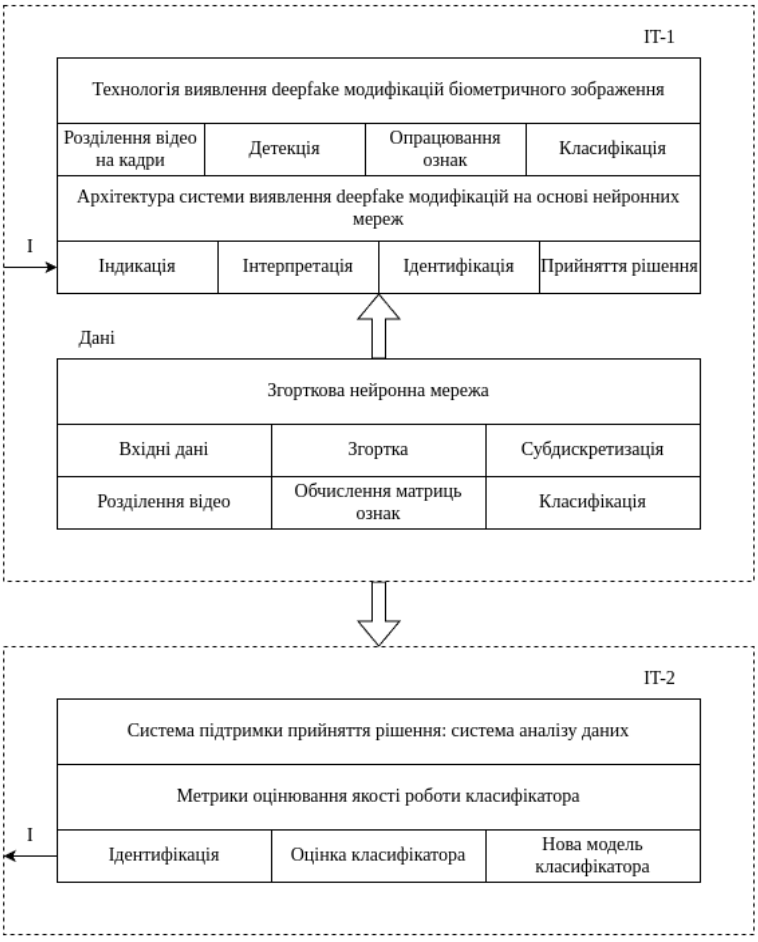

Рис. 1. Структура базового підходу до виявлення deepfake-модифікацій біометричного зображення

окремі функціональні модулі для підвищення ефективності та адаптивності алгоритму виявлення deepfake-модифікацій (рис. 2).

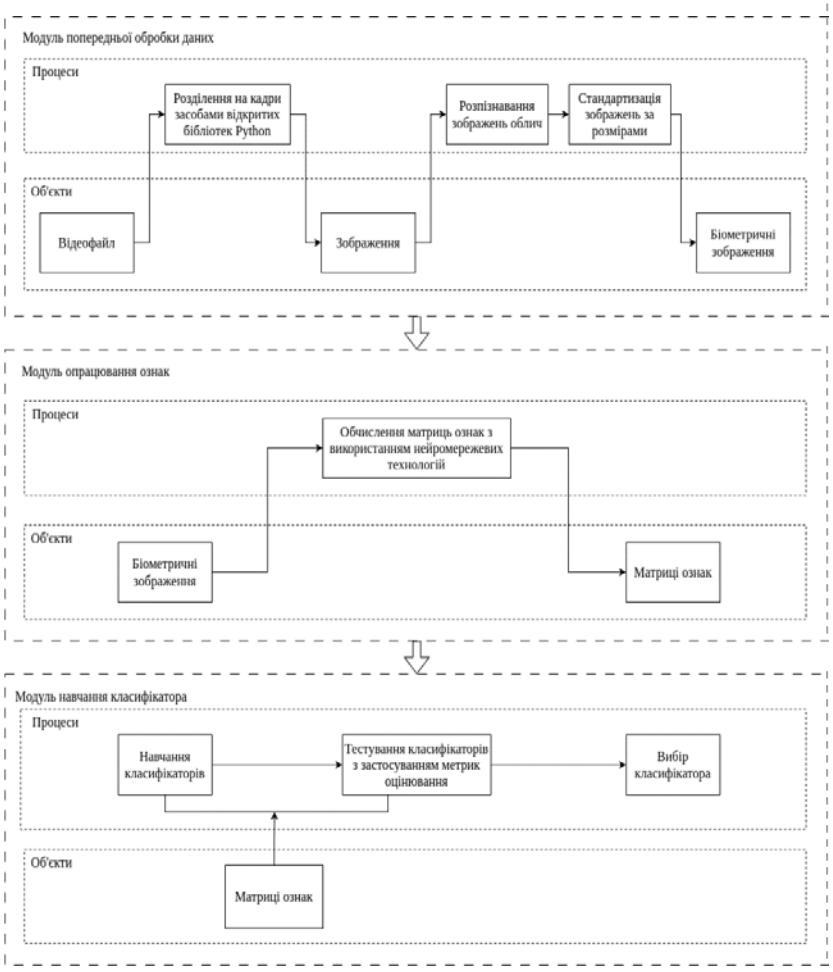

Рис. 2. Модульна архітектура системи виявлення deepfake-модифікацій

Модульна архітектура нейромережевої системи виявлення deepfake-модифікацій реалізує взаємозв'язаний алгоритм «попередня обробка даних – опрацювання ознак – навчання класифікатора», який функціонально розгорнутий згортковою нейронною мережею в просторі «вхідні дані – згортка – субдискретизація» і забезпечує «індикацію – інтерпретацію – ідентифікацію – прийняття рішення».

Модуль попередньої обробки даних системи виявлення deepfake-модифікацій функціонально реалізує алгоритм:

1) розділення відеофайлу на окремі кадри засобами відкритих бібліотек Python;

2) розпізнавання детектором біометричних зображень засобами нейронної мережі;

3) створення нових стандартизованих за розмірами біометричних зображень.

Модуль опрацювання ознак системи виявлення deepfake-модифікацій характерний алгоритмічною структурою: 1) відтворення нормалізованих біометричних зображень облич; оброблення зображень нейронною мережею; 2) обчислення матриць ознак засобами нейронної мережі і, на цій основі, побудова класифікатора зображень.

Модуль навчання класифікатора системи виявлення deepfake-модифікацій реалізує функціональний алгоритм: 1) навчання класифікатора; 2) тестування класифікатора на основі метрик оцінювання; 3) рішення про допуск класифікатора — модифіковане зображення; не модифіковане зображення.

Точність класифікації системи виявлення deepfake-модифікацій біометричних зображень враховує: 1) чутливість та специфічність класифікатора; 2) індекс Юдена [12], що визначає оптимальне порогове значення класифікації біометричних зображень; 3) інформовано класифіковані біометричні зображення.

Конструктивний алгоритм IT2 «ідентифікація – оцінка класифікатора – нова модель класифікатора» реалізується системою підтримки прийняття рішення в просторі аналізу даних та застосування метрик оцінювання: точності класифікатора, площі під кривою та логарифмічної функції втрат, що позиціонує різницю між прогнозованою ймовірністю належності елемента до певного класу та фактичною ймовірністю належності від класифікатора [13].

Методологія безпеки виявлення deepfake-модифікацій на основі багаторівневої моделі нейромережевої IT

За проведеним аналізом відомих підходів до безпечних технологій виявлення deepfake-модифікацій біометричних зображень запропоновано: 1) створення методології безпеки нейромережевих інформаційних технологій виявлення deepfake-модифікацій біометричних зображень обличчя людини у просторі безпечної інтелектуалізації об'єктів інфраструктури суспільства; 2) розвиток комплексної системи безпеки інформаційного процесу «фаза – операція – обробка» згідно з рівнями «розді-

лення відеофайлу на кадри — детекція, опрацювання ознак — оцінювання точності класифікатора зображень». Методологія безпеки нейромережевої ІТ виявлення deepfake-модифікацій (рис. 3) створена за принципами системного і синергетичного підходів та спрямована на забезпечення основних профілів безпеки — конфіденційності та цілісності біометричних зображень.

ІТ

Інформаційна технологія виявлення deepfake-модифікацій біометричного зображення					
ІР	ІС		ІП	ІМ	У
Біометричні зображення	Система виявлення deepfake-модифікацій на основі нейронних мереж	Система аналізу даних (Система підтримки прийняття рішення)	Фаза Операція Обробка	Комп'ютерні мережі Безпровідні мережі	ISO/IEC 27001:2022 Модель "Плануй-Виконуй-Перевіряй-Дій)

ІР

Біометричні зображення			Відеофайл	Піксель
База даних	База знань	База моделей	Масив інформації	Сховище даних
Загрози		Захист		
Цілеспрямовані	Випадкові	Апаратний	Програмний	

ІС

Конфіденційність Цілісність Доступність	**Багаторівнева багатоланкова модель безпеки ІС**				Витік Модифікація Знищення
	Зовнішній рівень	Апаратно-програмне забезпечення	Рівень інформаційних процесів		
	Загрози		Захист		
	Цілеспрямовані	Випадкові	Апаратний	Програмний	

ІП

Інформаційні процеси		
Фаза	Операція	Обробка
Загрози	Захист	
Цілеспрямовані Випадкові	Апаратний	Програмний

ІМ

Комп'ютерні та безпровідні мережі						
Загрози						
Сніфери пакетів	IP-спуфінг	DoS-атаки	Парольні атаки	Man-In-The-Middle	Атака на рівні застосунків	Вірусні атаки
Захист						
Антиспуфінг, Криптографія	Контроль доступу, Фільтрація	Анти-DoS, Обмеження обсягу трафіку	Криптографічна автентифікація	Криптографія	Web application firewalls, web gateway services	Антивірусне програмне забезпечення

У

Управління							
Життєвий цикл інформації в ІТ				Комплексна система безпеки ІТ			
Введення	Зберігання	Обробка	Обмін	ІР	ІС	ІП	ІМ
Загрози		Захист		Загрози		Захист	
Цілеспрямовані	Випадкові	Методи	Засоби	Цілеспрямовані	Випадкові	Методи	Засоби

Рис. 3. Структура методології безпеки багаторівневої нейромережевої ІТ

Системний підхід — принципи ієрархічності, структуризації, цілісності, які дають підстави для створення комплексної системи безпеки ІТ у просторі оптимального поєднання методологічного, технічного (апаратного), програмного, нормативного забезпечення безпечного функціонування життєвого циклу інформації в системі та алгоритму інформаційного процесу виявлення на рівні «фаза – операція – обробка». Синергетичний підхід — властивість емерджентності, що проявляє одну з граней цілісності захисту інформації в ІТ припускає наявність властивостей, що властиві комплексній системі безпеки ІТ в цілому, але не властиві її окремим елементам — комплексним системам безпеки ІР, ІС, ІП, ІМ, У. Ядром аналітичної структури безпечної нейромережевої інформаційної технології є система виявлення deepfake-модифікацій біометричних зображень на основі нейронних мереж та система аналізу даних, які програмно орієнтовані на цілісну реалізацію інформаційного процесу «розділення відео на кадри – виявлення deepfake – обробка ознак – оцінювання класифікації зображень» і, на цій основі, прийняття рішення про достатню точність класифікатора deepfake-модифікацій відповідно вибраній моделі з можливістю її оновлення. В таблиці подана комплексна система безпеки інформаційного процесу виявлення deepfake-модифікацій на рівні обробки біометричного зображення згідно з концепцією «об'єкт – загроза – захист».

Комплексна система безпеки виявлення deepfake-модифікацій на етапі обробки

Об'єкт: інформаційні процеси	Загрози		Захист інформації	
Обробка	Цілеспрямовані	Випадкові	Апаратний	Програмний
Автоматизована оцінка точності класифікатора	Витік, порушення конфіденційності, цілісності даних та моделей. Неавторизований доступ Зловмисне програмне забезпечення. Розподілені атаки на відмову в обслуговуванні (DDoS).	Збої/ нестабільність роботи технічних пристроїв. Помилки оператора. Невиправлені вразливості програмного забезпечення	Luna SA HSM. Luna SP. Luna XML	Encrypt Easy Suricata. Webroot DNS. Protection 1Password. BitLocker. Bitdefender. Antivirus
Виявлення deepfake-модифікацій в режимі реального часу	Маніпулювання даними. Інверсія моделі. Отруєння даних. Змагальні приклади Відмова в обслуговуванні	Технічні несправності мережі і компонентів. Вплив зовнішніх факторів. Пошкодження даних	nShield Connect HSM. Гряда-301. Бар'єр-301. Канал-301	ManageEngine. Log360. BitLocker
Послідовна обробка зображень	Отруєння даних: Приклади змагальності: Маніпуляція моделлю: Витік, порушення конфіденційності, цілісності даних та моделей. Злам криптографічного захисту	Збої в мережі. Фізичне пошкодження обладнання. Практики управління даними	Luna SA4 HSM. Luna PCM. Cisco Firepower. 2130 NGFW	Cisco UVPN-ZAS. BitLocker
Оцінювання точності класифікації з використанням хмарних обчислень	Витік, порушення конфіденційності, цілісності даних та моделей. Зловмисне програмне забезпечення. Розподілені атаки на відмову в обслуговуванні (DDoS).	Пошкодження даних. Збої в мережі. Невиправлені вразливості програмного забезпечення DoS на стороні постачальника послуг. Помилки оператора	Cisco Firepower. Palo Alto. Networks PA-7000 Series	Webroot DNS. Protection AlienVault USM

Сегмент міжнародних стандартів у сфері кібербезпеки: ISO/IEC 27034:2017, IEC 61508-3:2010, ISO/IEC 13335-1:2004 представляє нормативне забезпечення методології безпеки нейромережевої ІТ. Коаліція походження та автентичність вмісту C2PA опублікувала стандарт [14], який представляє методи підтвердження цифрового походження контенту. Стандарт визначає сценарії, робочі процеси та вимоги для методів підтвердження та перевірки інформації, що стосується створення та зміни медіафайлів. Застосування цих методів дозволяє редакторам контенту створювати захищені від фальсифікації носії, фіксуючи інформацію про те, хто створив або змінив цифровий кон-

тент і як він був змінений.

Висновки

В роботі запропоновано методологічні засади безпеки нейромережевих технологій у просторі проблеми виявлення deepfake-модифікацій біометричних зображень на основі: 1) базового підходу до виявлення deepfake-модифікацій; 2) структури методології безпеки нейромережевих багаторівневих інформаційних технологій; 3) комплексної системи безпеки інформаційного процесу виявлення deepfake-модифікацій на етапі оброблення згідно з концепцією «об'єкт – загроза – захист», що є розвитком системних підходів до безпечного виявлення deepfake-модифікацій засобами нейронних мереж.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

[1] H. Lasi, P. Fettke, H.G. Kemper, T. Feld, and M. Hoffmann, "Industry 4.0," *Business & Information Systems Engineering*, no. 6, pp. 239-242, 2014. https://doi.org/10.1007/s12599-014-0334-4 .

[2] *Стратегія кібербезпеки України* (2021—2025 року). [Електронний ресурс]. Режим доступу: https://www.rnbo.gov.ua/files/2021/STRATEGIYA%20KYBERBEZPEKI/proekt%20strategii_kyberbezpeki_Ukr.pdeepfake .

[3] Directorate-General for Communications Networks. Content and Technology, Cyber resilience act: new EU cybersecurity rules ensure more secure hardware and software products, *European Commission*, 2022. Available: https://data.europa.eu/doi/10.2759/543836 .

[4] Y. Shtefaniuk, and I. Opirskyy, "Comparative Analysis of the Efficiency of Modern Fake Detection Algorithms in Scope of Information Warfare," in *2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, pp. 207-211, 2021. https://doi.org/10.1109/IDAACS53288.2021.9660924.1 .

[5] Ю. Я. Бобало, В. Б. Дудикевич, і Г. В. Микитин. *Стратегічна безпека системи «об'єкт – інформаційна технологія»*. Львів, Україна: вид-во НУ «Львівська політехніка», 2020.

[6] M. Choraś, M. Pawlicki, D. Puchalski, and R. Kozik, "Machine Learning – The Results Are Not the only Thing that Matters! What About Security, Explainability and Fairness?" in *Lecture Notes in Computer Science*, vol. 12140. Springer, Cham. https://doi.org/10.1007/978-3-030-50423-6_46 .

[7] N. Papernot, P. McDaniel, A. Sinha, and M. P. Wellman, "SoK: Security and Privacy in Machine Learning," in *2018 IEEE European Symposium on Security and Privacy (EuroS&P)*, London, UK, 2018, pp. 399-414. https://doi.org/10.1109/EuroSP.2018.00035 .

[8] H. Bae, J. Jang, D. Jung, H. Jang, H. Ha, and S. Yoon. "Security and Privacy Issues in Deep Learning," 2018 ArXiv, abs/1807.11655.

[9] Q. Xu, M. Tanvir Arafin, and G. Qu, "Security of Neural Networks from Hardware Perspective: A Survey and Beyond," in *2021 26th Asia and South Pacific Design Automation Conference (ASP-DAC)*, Tokyo, Japan, 2021, pp. 449-454.

[10] X. Cao, and N. Z. Gong. "Understanding the Security of Deepfake Detection," in: *Digital Forensics and Cyber Crime. ICDF2C 2021. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering*, vol. 441. Springer, Cham, 2022.

[11] В. Б. Дудикевич, Г. В. Микитин, і Х. С. Руда, «Застосування глибокого навчання для виявлення deepfake-модифікацій біометричного зображення," *Сучасна спеціальна техніка*, № 1, с. 13-22, 2022.

[12] W. J. Youden, "Index for rating diagnostic tests," *Cancer*, no. 3, pp. 32-35, 1950. https://doi.org/ 10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3. PMID 15405679.

[13] E. Altuncu, V. Franqueira, and L. Shujun, *Deepfake: Definitions, Performance Metrics and Standards, Datasets and Benchmarks, and a Meta-Review*, 2022. 10.48550/arXiv.2208.10913 .

[14] *C2PA. 2020. Coalition for Content Provenance and Authenticity*. [Electronic resource]. Available: https://c2pa.org/ .

Рекомендована кафедрою менеджменту та безпеки інформаційних систем ВНТУ

Стаття надійшла до редакції 13.02.2024

Микитин Галина Василівна — д-р техн. наук, професор, професор кафедри захисту інформації, email: halyna.v.mykytyn@lpnu.ua ;

Руда Христина Степанівна — аспірантка кафедри захисту інформації, email: khrystyma.s.ruda@lpnu.ua .

Національний університет «Львівська політехніка», Львів;

Яремчук Юрій Євгенович — д-р техн. наук, професор, директор центру інформаційних технологій і захисту інформації, професор кафедри менеджменту та безпеки інформаційних систем, email: yurevyar@vntu.edu.ua .

Вінницький національний технічний університет, Вінниця

H. V. Mykytyn[1]
Kh. S. Ruda[1]
Yu. Ye. Yaremchuk[2]

Security Methodology of Neural Network-Based Information Technologies for Detection of Deepfake-Modifications of Biometric Image

[1]Lviv Polytechnic National University;
[2]Vinnytsia National Technical University

One of the functional vectors of the Cybersecurity Strategy of Ukraine is the development and implementation of protection systems for various information platforms in society's infrastructure, particularly focusing on creating safe technologies to detect deepfake-modifications of biometric images, based on neural networks in cyberspace. This paper presents the security principles of neural network information technologies (IT) within the context of deepfake-modifications. It delineates a basic approach for safely detecting deepfake-modifications in biometric images and outlines a security methodology for multi-level neural network IT systems, organized according to the "object – threat – protection" concept. The basic approach integrates information neural network technology with decision support IT, structured by a modular architecture for detecting deepfake modifications. This architecture operates across the stages of "pre-processing of feature data – classifier training." The core of the IT security methodology emphasizes the integrity of neural network systems for detecting deepfake-modifications in biometric images, coupled with data analysis systems that execute the information process of "dividing video files into frames – detecting and processing features – evaluating the accuracy of image classifiers. The security methodology for multi-level neural network IT relies on systemic and synergistic approaches to construct a comprehensive IT security system. This system accounts for the possibility of emergent threats and incorporates cutting-edge countermeasure technologies at both hardware and software levels. The proposed comprehensive security system for detecting deepfake-modifications in biometric images encompasses hardware and software tools across several segments: automated classifier accuracy assessment, real-time deepfake-modification detection, sequential image processing, and classification accuracy evaluation utilizing cloud computing.

Keywords: intellectualization, cyber security, biometric image, deepfake, information technology, neural networks, detection system, basic approach, security methodology, comprehensive security system.

Mykytyn Halyna V. — Dr. Sc. (Eng.), Professor, Professor of the Chair of Information Security, halyna.v.mykytyn@lpnu.ua ;
Ruda Khrystyna S. — Post-Graduate Student of the Chair of Information Security, email: khrystyna.s.ruda@lpnu.ua ;
Yaremchuk Yurii Ye. — Dr. Sc. (Eng.), Professor, Director of the Center of Information Technologies and Information Security, Professor of the Chair of Management and Information Security, email: yurevyar@vntu.edu.ua