УДК 004.89:004.8 **DOI: https://doi.org/10.30837/PMW.2026.T2.376**

КОНЦЕПЦІЯ АРХІТЕКТУРИ ІНТЕЛЕКТУАЛЬНОЇ МУЛЬТИМЕДІЙНОЇ ПЛАТФОРМИ ДЛЯ РОЗПІЗНАВАННЯ ФЕЙКОВОГО КОНТЕНТУ З УРАХУВАННЯМ UI/UX-ПІДХОДІВ

Марків О.О.
к.т.н., доцент, кафедра Інформаційних систем та мереж,
НУ «Львівська політехніка»
ORCID ID: 0000-0002-1691-1357
Висоцька В.А.
д.т.н., доцент, кафедра Інформаційних систем та мереж,
НУ «Львівська політехніка»
ORCID ID: 0000-0001-6417-3689
Лозинська О.В.
к.т.н., доцент, кафедра Інформаційних систем та мереж,
НУ «Львівська політехніка»
ORCID ID: 0000-0002-5079-0544

Анотація. *У розділі досліджено концептуальні засади побудови інтелектуальної мультимедійної платформи для розпізнавання фейкового контенту в цифровому інформаційному середовищі з урахуванням сучасних UI/UX-підходів. Проаналізовано сучасні методи автоматизованого виявлення фейкового контенту із застосуванням технологій штучного інтелекту, машинного навчання, обробки природної мови (NLP). Запропоновано концепцію архітектури інтелектуальної мультимедійної платформи з інтеграцією UI/UX-підходів у процес проєктування платформи.*

Ключові слова: *фейковий контент, машинне навчання, опрацювання природної мови, мультимедійний контент, веб-інтерфейс, UI/UX-підходи, дезінформація.*

Вступ

Актуальність дослідження зумовлена стрімким поширенням мультимедійного контенту, дезінформації, deepfake-технологій та фейкових інформаційних повідомлень у відкритих веб-ресурсах і соціальних медіа, що створює суттєві загрози інформаційній безпеці суспільства та цифровій довірі користувачів.

Розробка інтелектуальної мультимедійної платформи допомагає автоматизувати розпізнавання дезінформації, надає користувачам зручні інтерфейси для перевірки достовірності, забезпечує масштабованість та адаптивність системи до нових типів фейків.

У межах дослідження запропоновано концепцію веб-орієнтованого програмного прототипу платформи для виявлення фейкових новин і маніпулятивного контенту. Розроблений прототип передбачає аналіз текстових повідомлень та графічного контенту на основі спеціалізованих датасетів

дезінформації. Особливістю системи є надання користувачу пояснення результатів аналізу за допомогою засобів візуалізації, а також реалізація елементів навчального середовища для формування навичок розпізнавання інформаційних маніпуляцій.

Структура запропонованої системи включає низку функціональних модулів, кожен із яких виконує окремі завдання в процесі аналізу та інтерпретації контенту. Зокрема, модуль аналізу тексту забезпечує визначення ймовірності належності повідомлення до фейкового контенту із застосуванням методів обробки природної мови. Для реалізації цього модуля передбачається використання підходу TF-IDF для векторизації текстових даних та алгоритму логістичної регресії для здійснення класифікації. Результатом роботи моделі є оцінка ймовірності фейковості інформаційного повідомлення.

Модуль виявлення маніпулятивних конструкцій орієнтований на ідентифікацію підозрілих мовних елементів, зокрема, емоційно забарвлених слів, узагальнень та сенсаційних формулювань. Реалізація такого функціоналу можлива шляхом поєднання rule-based підходів та інструментів NLP-аналізу. Додатково система передбачає модуль класифікації типів маніпулятивного впливу, який визначає характер інформаційного впливу, наприклад, емоційний тиск, узагальнення або “викривлення” фактів.

Запропонована система передбачає застосування поліграфічних засобів комунікації, зокрема розроблення інформаційних матеріалів та інфографіки, що містять основні ознаки фейкових новин і маніпулятивного контенту. Інтеграція QR-кодів забезпечує швидкий перехід користувачів до веб-платформи, що дозволяє поєднати офлайн- та онлайн-середовище взаємодії.

Таким чином, запропонований прототип платформи забезпечує не лише автоматизоване виявлення фейкового контенту, але й пояснення механізмів інформаційного впливу, що є важливим чинником підвищення довіри користувачів до системи та розвитку критичного мислення. Поєднання вебтехнологій, методів машинного навчання, засобів Explainable AI та поліграфічних рішень формує комплексний підхід до протидії інформаційним маніпуляціям у сучасному цифровому середовищі.

Мета та задачі дослідження

Метою дослідження є розроблення концепції та програмного прототипу веб-орієнтованої платформи для автоматизованого виявлення фейкових новин і маніпулятивного контенту з використанням методів машинного навчання, обробки природної мови, а також створення інструментів візуалізації результатів аналізу та навчального середовища для розвитку навичок критичного сприйняття інформації.

Для досягнення поставленої мети необхідно виконати такі завдання:

– проаналізувати сучасний стан проблеми поширення дезінформації, фейкових новин та маніпулятивного контенту у цифровому середовищі;

– дослідити існуючі методи та підходи до автоматичного виявлення фейкової інформації із застосуванням технологій машинного навчання та NLP;

– розробити архітектуру веб-орієнтованої платформи для аналізу та класифікації інформаційного контенту;

– реалізувати модуль аналізу текстових даних;

– запропонувати підхід до виявлення маніпулятивних мовних конструкцій із використанням rule-based методів та NLP-аналізу;

– забезпечити візуалізацію результатів аналізу та інтеграцію офлайн- і онлайн-інформаційних матеріалів із використанням, наприклад, QR-кодів.

Основна частина

Аналіз літературних джерел та практичних рішень

Аналіз наукових джерел показує, що сучасні дослідження у сфері розпізнавання фейкового контенту концентруються переважно на використанні моделей глибокого навчання. Нижче наведено узагальнену аналітичну таблицю сучасних наукових праць, дотичних до теми архітектури інтелектуальних мультимедійних платформ для розпізнавання фейкового контенту з урахуванням UI/UX-підходів (табл.1).

Таблиця 1 – Аналіз сучасних наукових праць

Автори / Рік	**Країна**	**Тема дослідження**	**Методи / Технології**	**Основні результати**	**UX/UI-аспекти**
Agarwal et al., 2020	Великобританія, University of Cambridge	Detecting AI-manipulated Media	Multimodal Transformers	Підвищення точності мультимодального аналізу	Human-in-the-loop UX, пояснювані результати
Zhou & Zafarani, 2020	Німеччина/США	Multimodal Fake News Detection	CNN + Transformer (text+image)	Комбінований аналіз збільшує точність	Dashboards, кольорові індикатори достовірності
Ángel Fernández Gambín та ін., 2021	Іспанія, Universidad de Vigo	*Deep Fakes Disentangling Terms in the Proposed EU Artificial Intelligence*	Deep Learning, Big Data, multimedia forensics	Аналіз AI-generated відео та медіа-маніпуляцій	Пропозиції щодо візуалізації ризиків фейку
Stroebel, L., Llewellyn, M., Hartley, T., Ip, T. S., & Ahmed, M. (2023)	Велика Британія	*A Systematic Literature Review on the Effectiveness of Deepfake Detection Techniques.*	*Convolutional Neural Networks, LSTM, Transformer-моделі*	*Найефективнішими є: hybrid approaches; multimodal systems; deep learning;forensic analysis*	*з аналізу підходів до deepfake detection можна виділити практичні UX/UI-аспекти, наприклад, Explainability UI*

Продовження таблиці 1

Автори / Рік	**Країна**	**Тема дослідження**	**Методи / Технології**	**Основні результати**	**UX/UI-аспекти**
Микитин, Г. В., Руда, Х. С., & Яремчук, Ю. Є. (2024)	Україна	*Методологія безпеки нейромережевих інформаційних технологій виявлення deepfake-модифікацій біометричного зображення*	методологія захисту нейромережевих систем виявлення deepfake-маніпуляцій біометричних зображень	Комплексна система, що поєднує обробку зображень, детекцію ознак і оцінювання точності класифікації	візуалізація аналізу, зниження когнітивного навантаження

Розроблення та реалізація прототипу концептуальної архітектури мультимедійної платформи для виявлення фейкових відгуків

Архітектура реалізована за багаторівневим принципом і включає підсистеми взаємодії з користувачем, серверної обробки, аналізу контенту та зберігання даних (рис. 1).

Такий підхід забезпечує модульність і можливість інтеграції різних методів перевірки мультимедійної інформації.

Верхній рівень архітектури представлений *модулем взаємодії з користувачем (UI/UX)*, який включає веб-інтерфейс, мобільний застосунок та індикатор достовірності. Модуль UI/UX забезпечує взаємодію користувача з мультимедійною платформою та відповідає за подання результатів аналізу у зрозумілому вигляді. Основною метою модуля є забезпечення доступності функціоналу системи через різні клієнтські інтерфейси.

Веб-інтерфейс реалізує браузерний доступ до платформи та надає користувачу можливість:

– завантаження мультимедійного контенту;

– перегляду результатів перевірки;

– отримання пояснень щодо достовірності.

Мобільний застосунок забезпечує доступ до функціоналу системи з мобільних пристроїв та забезпечує оперативну перевірку контенту, можливість push-сповіщення, синхронізацію результатів, локальне кешування даних.

Індикатор достовірності є *модулем візуалізації результатів аналізу* та відображає рівень довіри до контенту, ризик фальсифікації, категорію загрози, короткий висновок системи. Індикатор може реалізовуватись у вигляді шкали, кольорового маркера або числового коефіцієнта.

Backend-модуль є центральною серверною частиною системи та забезпечує обробку запитів, маршрутизацію даних і координацію роботи всіх сервісів, зокрема, сервіс логування відповідає за реєстрацію подій системи, ведення журналів активності, моніторинг помилок, аналіз продуктивності та ін.

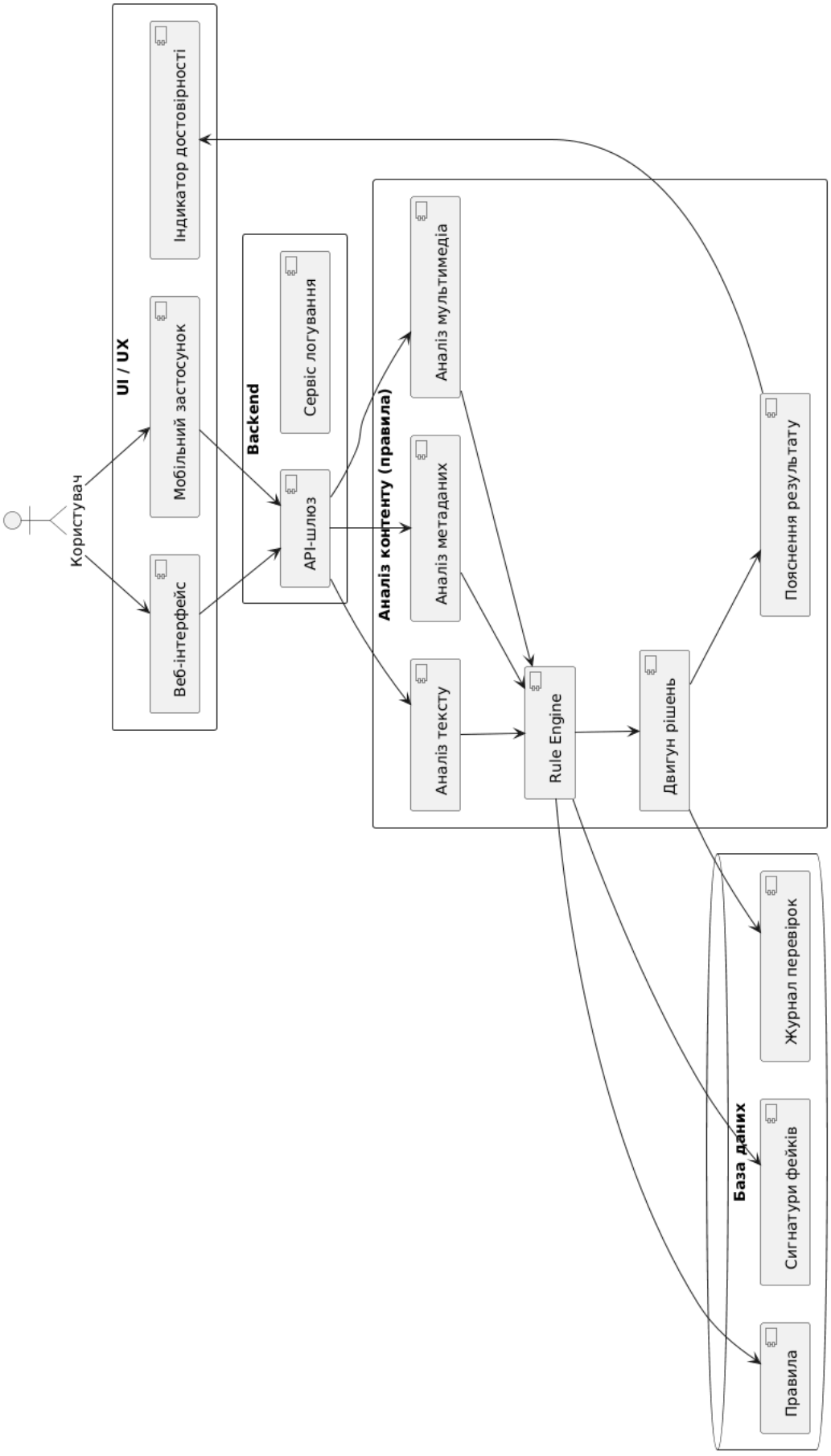

Рисунок 1 – Архітектура мультимедійної платформи для виявлення фейкового контенту

Ключовим елементом системи є *модуль аналізу контенту*, побудований на основі правил та евристичних моделей, виконує обробку текстових даних із використанням методів лінгвістичного аналізу, виявлення маніпулятивних конструкцій і семантичної оцінки.

Модуль аналізу метаданих досліджує службову інформацію файлів, часові мітки, геолокаційні параметри та інші атрибути цифрових об'єктів. Аналіз метаданих дозволяє виявляти ознаки модифікації контенту. Модуль аналізу мультимедіа здійснює перевірку зображень, аудіо- та відеофайлів на предмет монтажу, генерації штучним інтелектом або інших ознак фальсифікації, пошук deepfake-контенту, аналіз шумів та артефактів, перевірка цілісності медіаданих, спектральний аналіз аудіо.

Rule Engine є центральним компонентом логічної обробки, передбачає агрегацію результатів аналізу, застосування правил перевірки, оцінювання ризиків, формування проміжних висновків. Механізм працює на основі набору формалізованих правил та сигнатур фальсифікацій.

Двигун рішень виконує остаточне оцінювання достовірності, передбачає обчислення рівня довіри, класифікацію загроз, прийняття рішення щодо фейковості, генерацію статусу перевірки. Для цього пропонується використовувати експертні системи, байєсівські моделі, ML-класифікатори, системи оцінювання ризику.

Модуль пояснення результату відповідає за формування інтерпретованих висновків для користувача, передбачає пояснення причин оцінки відображення знайдених аномалій, формування рекомендацій, забезпечення прозорості роботи системи. Модуль підвищує рівень довіри користувача до автоматизованої перевірки.

Модуль бази даних та сигнатури фейків забезпечує централізоване зберігання інформації системи, містить шаблони маніпуляцій, відомі deepfake-маркери, характеристики фальсифікованого контенту.

Особливості використання UI/UX-підходів

У сучасних інтелектуальних системах UI/UX-дизайн відіграє не менш важливу роль, ніж алгоритмічна складова аналізу даних. Результати перевірки повинні бути не лише точними, але й зрозумілими для користувача:

– принцип usability передбачає створення інтуїтивно зрозумілого інтерфейсу з мінімальним когнітивним навантаженням на користувача.

– принцип accessibility забезпечує доступність платформи для користувачів із різними фізичними можливостями та типами пристроїв.

– принцип adaptive design дозволяє коректно відображати інформацію на веб- та мобільних платформах.

– принцип trust UX орієнтований на формування довіри до автоматизованих рішень з використанням, наприклад, кольорових індикаторів ризику.

Таким чином, UI/UX-підходи стають невід’ємною складовою сучасних платформ автоматизованого аналізу цифрового контенту.

Приклад MVP системи виявлення фейкових відгуків із врахуванням бальної системи оцінювання(score-based) має наступні особливості.

1. Використання наперед створеного датасету дезінформації (рис. 2).

```
data = [
    ["ЦЕ НАЙКРАЩИЙ СЕРВІС!!! 🔥🔥🔥", "2026-04-30", 1],
    ["Дуже рекомендую всім", "2026-04-29", 10],
    ["Супер сервіс, все добре", "2026-04-28", 3],
    ["ок ок ок ок", "2026-05-01", 1],
    ["Найкращий найкращий сервіс", "2026-05-01", 2],
]
```

index	text	score	label	reasons
0	ЦЕ НАЙКРАЩИЙ СЕРВІС!!! 🔥🔥🔥	5	SUSPICIOUS	Короткий текст,Емоційність (CAPS/!!!),Новий акаунт
1	Дуже рекомендую всім	2	NORMAL	Короткий текст
2	Супер сервіс, все добре	2	NORMAL	Короткий текст
3	ЖАХ!!! УЖАС!!!	5	SUSPICIOUS	Короткий текст,Емоційність (CAPS/!!!),Новий акаунт
4	ок ок ок ок	6	SUSPICIOUS	Короткий текст,Повторювані слова,Новий акаунт
5	Найкращий найкращий сервіс	6	SUSPICIOUS	Короткий текст,Повторювані слова,Новий акаунт
6	Все нормально, дякую	2	NORMAL	Короткий текст
7	РЕКОМЕНДУЮ ВСІМ!!!	5	SUSPICIOUS	Короткий текст,Емоційність (CAPS/!!!),Новий акаунт

Рисунок 2 – Приклад датасету дезінформації

2. Правила оцінювання за шкалою балів за певними ознаками.

```
def is_short(text):
    return len(text.split()) < 10

def caps_or_exclamations(text):
    return bool(re.search(r"[A-ZA-ЯЇІЄ]{5,}|!{2,}", text))

def repeated_words(text):
    words = text.lower().split()
    return len(words) != len(set(words))

def clean_text(text):
    return re.sub(r"[^a-zA-Za-яA-Я0-9 ]", "", text.lower())
```

З чітким зазначенням додавання чи віднімання балів.

```
    # CAPS / !!!
    if caps_or_exclamations(text):
        score += 1
        reasons.append("Емоційність (CAPS/!!!)")

    # повтори слів
    if repeated_words(text):
        score += 2
        reasons.append("Повторювані слова")
```

```
# нові або активні акаунти
if df.loc[i, "user_reviews_count"] < 3:
    score += 2
    reasons.append("Новий аккаунт / недостатньо історії")
```

З чіткими прописаними правилами отримання результатів (таблиця 2).

```
if score >= 7:
    label = "FAKE"
elif score >= 4:
    label = "SUSPICIOUS"
else:
    label = "NORMAL"
```

Таблиця 2 – Правила отримання результатів

text	score	label	reasons
"ЦЕ НАЙКРАЩИЙ СЕРВІС!!!"	6	SUSPICIOUS	емоційність, короткий, новий акаунт
"ок ок ок ок"	7	FAKE	повтори, короткий, новий акаунт
"Все нормально, дякую"	0	NORMAL	—

Запропонований MVP системи виявлення фейкових відгуків реалізує score-based підхід, у якому аналіз тексту виконується за допомогою набору евристичних правил та бальної системи оцінювання. Система використовує попередньо сформований датасет і перевіряє відгуки на наявність підозрілих ознак, зокрема емоційності тексту, повторення слів, коротких повідомлень та низької активності користувача. Такий підхід забезпечує прозорість логіки прийняття рішень і може слугувати основою для подальшого впровадження методів машинного навчання та NLP-аналізу.

Висновки

У результаті проведеного дослідження встановлено, що проблема поширення фейкового контенту та deepfake-технологій є одним із ключових викликів сучасного цифрового середовища. Аналіз міжнародних та українських наукових праць показав активний розвиток методів автоматизованого виявлення дезінформації та необхідність покращення ефективності, враховуючи взаємодію з користувачем, прозорість результатів аналізу та адаптивність інтерфейсів.

Запропоновано концепцію веб-орієнтованої мультимедійної платформи для автоматизованого виявлення фейкового контенту. Розроблена архітектура системи реалізована за багаторівневим принципом та включає модулі взаємодії з користувачем, серверної обробки, аналізу контенту, Rule Engine, двигуна рішень, модуля пояснення результатів і бази даних сигнатур фейків.

Окрему увагу приділено питанням UI/UX-дизайну та візуалізації результатів аналізу. Запропоновано використання індикаторів достовірності, адаптивного дизайну, кольорових маркерів ризику та пояснювальних механізмів,

що підвищують рівень довіри користувачів до системи. Важливою складовою платформи є інтеграція навчальних елементів та поліграфічних матеріалів із QR-кодами, які поєднують офлайн- та онлайн-середовище й сприяють розвитку навичок критичного сприйняття інформації.

Перспективи подальших досліджень полягають у розширенні функціональності системи за рахунок використання глибоких нейронних мереж, а також удосконалення механізмів автоматичного навчання та персоналізації взаємодії з користувачем. Це дозволить підвищити точність виявлення фейкового контенту та адаптивність системи до нових типів інформаційних маніпуляцій у сучасному цифровому середовищі.

Результати дослідження можуть бути використані під час розроблення інтелектуальних систем інформаційної безпеки, платформ фактчекінгу, мультимедійних аналітичних сервісів, а також у наукових і прикладних дослідженнях у галузях штучного інтелекту, цифрових комунікацій, UX/UI-дизайну та кібербезпеки.

Подяка.

Дослідження було проведено за грантової підтримки Міністерства освіти і науки України «Методи та засоби виявлення дезінформації у соціальних мережах на основі технологій глибинного навчання» в рамках проєкту № 0125U001852.

Список літератури.

1. Kaliyar, R. K., Goswami, A., Narang, P., & Sinha, S. (2021). FakeBERT: Fake News Detection in Social Media with a BERT-Based Deep Learning Approach. Multimedia Tools and Applications, 80, 11765-11788. https://doi.org/10.1007/s11042-020-10183-2.

2. Agarwal, S., Farid, H., Gu, Y., He, M., Nagano, K., & Li, H. (2020). Detecting Deep-Fake Videos from Appearance and Behavior. IEEE International Workshop on Information Forensics and Security (WIFS). https://doi.org/10.1109/WIFS49906.2020.9360891.

3. Zhou, X., Wu, J., & Zafarani, R. (2020). SAFE: Similarity-Aware Multi-Modal Fake News Detection. Advances in Knowledge Discovery and Data Mining (PAKDD 2020). (p. 354-367). Springer. https://doi.org/10.1007/978-3-030-47436-2_27.

4. Fernández, Á. (2021). "Deep Fakes": Disentangling Terms in the Proposed EU Artificial Intelligence Act. UFITA – Archiv für Medienrecht und Medienwissenschaft, 85(2), 392-433. https://doi.org/10.5771/2568-9185-2021-2-392.

5. Ponce, A., & Ponce Rodriguez, R.A. (2020). An Analysis of the Supply of Open Government Data. Future Internet, 12(11), 186. https://doi.org/10.3390/fi12110186.

6. Stroebel, L., Llewellyn, M., Hartley, T., Ip, T.S., & Ahmed, M. (2023). A Systematic Literature Review on the Effectiveness of Deepfake Detection Techniques. Journal of Cyber Security Technology, 7(2), 83-113. https://doi.org/10.1080/23742917.2023.2192888.

7. Lotfi, S., Mirzarezaee, M., Hosseinzadeh, M., & Seydi, V. (2021). Detection of Rumor Conversations in Twitter Using Graph Convolutional Networks. Applied Intelligence, 51(7), 4774-4787. https://doi.org/10.1007/s10489-020-02036-0.

8. Babaei, R., Cheng, S., Duan, R., & Zhao, S. (2025). Generative Artificial Intelligence and the Evolving Challenge of Deepfake Detection: A Systematic Analysis. Journal of Sensor and Actuator Networks, 14(1), 17. https://doi.org/10.3390/jsan14010017.

9. Mittal, T., Bhattacharya, U., Chandra, R., Bera, A., & Manocha, D. (2020). Emotions Don’t Lie: An Audio-Visual Deepfake Detection Method Using Affective Cues. 28th ACM International Conference on Multimedia (ACM MM). (p. 2823-2832). https://doi.org/10.1145/3394171.3413570.

10. Fernández Gambín, Á., Yazidi, A., Vasilakos, A., Haugerud, H., & Djenouri, Y. (2024). Deepfakes: current and future trends. Artificial Intelligence Review, 57, Article 64. https://doi.org/10.1007/s10462-023-10679-x.

11. Микитин, Г.В., Руда, Х.С., & Яремчук, Ю.Є. (2024). Методологія безпеки нейромережевих інформаційних технологій виявлення deepfake-модифікацій біометричного зображення. Вісник Вінницького політехнічного інституту, 172(1), 74-80. https://doi.org/10.31649/1997-9266-2024-172-1-74-80.

12. Микитин, Г.В., & Руда, Х.С. (2024). Conceptual approach to detecting deepfake modifications of biometric images using neural networks. Computer Systems and Networks. https://doi.org/10.23939/csn2024.01.124.