DOI: 10.18372/2310-5461.65.19758
УДК 004.738.5

***В. В. Носов***, канд. техн. наук, доцент
Харківський національний університет внутрішніх справ
https://orcid.org/0000-0002-7848-6448
e-mail: vitnos@ukr.net;

***О. В. Манжай***, канд. юр. наук, професор
Харківський національний університет внутрішніх справ
http://orcid.org/0000-0001-5435-5921
e-mail: sofist@ukr.net;

***С. Д. Лучик***, д-р екон. наук, професор
Харківський національний університет внутрішніх справ
https://orcid.org/0000-0003-0757-1140
e-mail: luchiksvitlana@gmail.com;

***В. Є. Лучик***, д-р екон. наук, професор,
Харківський національний університет внутрішніх справ
https://orcid.org/0000-0001-6007-910X
e-mail: luchik-vasil@ukr.net;

***В. О. Товстик,***
Харківський національний університет внутрішніх справ
e-mail: vadymka38@gmail.com

# ПРАКТИЧНА ОЦІНКА ЗАСОБІВ ВИЯВЛЕННЯ ІНФОРМАЦІЙНОГО ВПЛИВУ НА КОРИСТУВАЧІВ ГЛОБАЛЬНОЇ МЕРЕЖІ INTERNET

**Вступ**

Поєднуючи мільярди людей у всьому світі, Інтернет є основою сучасного інформаційного суспільства. Станом на жовтень 2024 року у світі налічувалося 5,52 мільярда користувачів Інтернету, що становило 67,5 відсотка світового населення. З них 5,22 мільярда, або 63,8 відсотка населення світу, були користувачами соціальних мереж [1]. Доступ людей до надійної та достовірної інформації став важливим етапом для виживання, прогресу та миру.

Проте, внаслідок постійного збільшення потенціалу інтернет-середовища та надання користувачам практично необмежених можливостей в отриманні інформації, виникає чимало загроз та ризиків для суспільства, які можуть мати довгострокові негативні наслідки. Зокрема, інформаційний вплив через Інтернет часто використовується для досягнення таких стратегічних цілей, як послаблення політичної стабільності держави, підрив довіри до інституцій або маніпулювання суспільною думкою. Актуальним прикладом є російські центри інформаційно-психологічних спеціальних операцій (ІПСО), які проводять широкомасштабні операції, що поєднують дезінформацію, кібератаки та психологічний вплив. Активно проводяться кампанії з дискредитації ключових осіб або подій через соціальні медіа та новинні платформи, зокрема, за допомогою так званих «фабрик тролів».

**Постановка проблеми**

Протистояння України військовій агресії РФ потребує постійно відстежувати інформаційний вплив на українських громадян, користувачів глобальної мережі Internet. Для виявлення дезінформації, фейків, кіберзлочинів та інших інформаційних загроз розроблена значна кількість методів та інструментів, більшість з яких залишаються невідомими широкому колу користувачів. Тому актуальною задачею статті є систематизація існуючих методів і засобів виявлення інформаційного впливу на інтернет-користувачів, проведення оцінювання їх ефективності та надання рекомендації щодо практичного їх використання з метою протидії та нейтралізації інформаційних загроз.

**Аналіз останніх досліджень і публікацій**

Враховуючи різноманітність видів інформаційного впливу на постійно зростаючу кількість користувачів глобальної мережі Інтернет та необхідність постійно протидіяти інформаційним загрозам, що виникають, питання виявлення інфор-

маційного впливу на користувачів Інтернету завжди актуальні і у науковій літературі накопичено чимало публікацій з розгляду цих питань.

В. Бурячок, А. Дудатьєв, В. Лужецький, В. Толубко, С. Толюпа, В. Хорошко у своїх працях розглядали концептуальні засади інформаційного впливу з визначенням об'єктів, ресурсів та процесів, що відбуваються у системі інформаційного впливу, досліджували моделі системи інформаційного впливу на соціальну частину соціотехнічної системи. На думку авторів, інформаційна безпека таких соціотехнічних систем, як держава, у великій мірі залежить від рівня захищеності соціуму [2].

Дослідження ризиків та потенціалу інтернет-середовища, яке є платформою для розповсюдження різноманітної інформації (дезінформації), що може загрожувати національній безпеці, здійснювали такі науковці як А. Аносов, О. Горун, Н. Дяків, О. Звоздецька, Н. Литвин, З. Пужняк, А. Ярош. Зокрема, автори вважають, що «дезінформація може призводити до розколу суспільства, підірвання єдності та створення напруженості між різними групами громадян, що може бути особливо небезпечним в умовах воєнного стану; часті випадки дезінформації можуть призводити до загальної недовіри будь-якій інформації, включаючи ту, яка є правдивою та достовірною, що може створити серйозні проблеми в передачі інформації громадянам; дезінформація може впливати на прийняття рішень на різних рівнях – від громадянського до державного, між тим погано підготовлені або маніпульовані дані можуть спричинити неправильне прийняття рішень та стратегії, що загрожуватиме національній безпеці» [3].

Моделі інформаційного протиборства як суперництва в інформаційній сфері з метою впливу на різні аспекти суспільних відносин є надзвичайно важливими, особливо в період гібридної війни, тому досліджувалися рядом вчених, серед яких О. Войтович, О. Гладка, А. Дудатьєв, І. Карпович, Л. Куперштейн, А. Тимракевич. Для визначення особливостей розповсюдження інформаційних загроз в умовах інформаційно-психологічного протиборства та знаходження оптимальних режимів поширення інформації і засобів нейтралізації інформаційних загроз науковці використовують методи математичного моделювання з урахуванням сучасних досягнень соціальної психології [4].

Безпосередньо психологічні механізми інформаційного впливу на особистість, суспільство, державу в цілому досліджували Алещенко В., Анісімович-Шевчук О., Белікова Т., Гречка С., Зінич М. Автори стверджують, що «інформаційно-психологічна безпека особистості покликана захистити особистість і суспільство від деструктивного інформаційного впливу» [5].

Попри широкий інтерес науковців до теми дослідження, докладного розгляду потребують питання практичної оцінки ефективності методів і засобів виявлення інформаційного впливу на користувачів глобальної мережі Інтернет.

***Метою*** дослідження є практична оцінка ефективності доступних засобів виявлення інформаційного впливу на користувачів глобальної мережі. Для досягнення поставленої мети необхідно вирішити низку завдань:

– ідентифікація методів та засобів виявлення інформаційного впливу на користувачів глобальної мережі Internet;

– розробка сценарію практичної оцінки окремих засобів виявлення інформаційного впливу;

– здійснення за розробленим сценарієм практичної оцінки визначених засобів виявлення інформаційного впливу.

**Виклад основного матеріалу**

Ґрунтовний та системний аналіз наукових публікацій з питань оцінювання контенту в глобальному інформаційному просторі дозволив виділити та згрупувати основні методи і засоби ідентифікації інформаційних впливів з метою оцінки їх потенційних ризиків для суспільства (рис. 1).

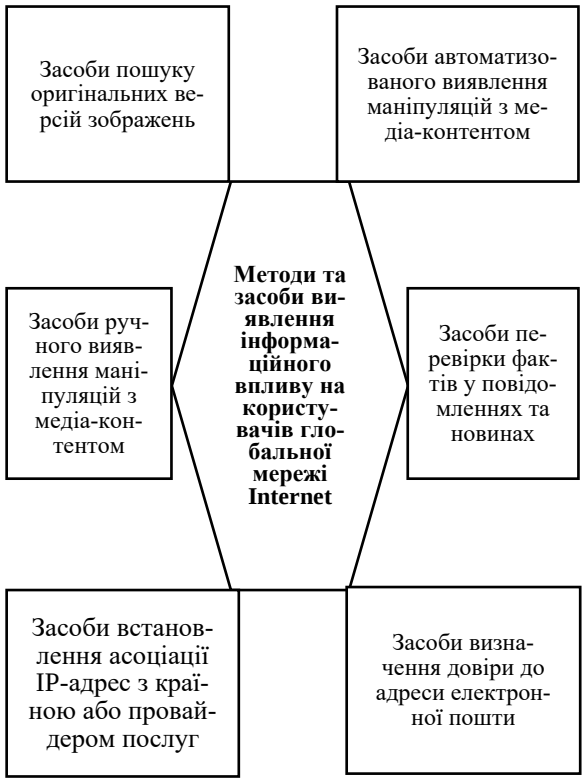


Рис. 1. Класифікація методів та засобів виявлення інформаційного впливу на користувачів глобальної мережі Internet

До першої групи, яку формують засоби перевірки фактів у повідомленнях та новинах (так званий фактчекінг), нами було віднесено такі інструменти як: FactBot by Snopes [6], PolitiFact [7],

FactCheck.org [8], ChatGPT [9]. Охарактеризуємо їх детальніше.

PolitiFact є інструментом для перевірки фактів, що аналізує висловлювання політиків, лідерів та ЗМІ, оцінюючи їх правдивість за допомогою шкали "Truth-O-Meter". Основні функції такі: «Перевірка фактів» – аналіз і оцінка публічних заяв на правдивість; «Пошук і доступ до архіву» – база перевірених тверджень за ключовими словами; «Освітні матеріали» – пояснення складних тем і фактів для громадськості.

FactBot by Snopes є інструментом, який допомагає користувачам перевіряти факти безпосередньо у соціальних мережах або месенджерах, тобто швидко відповідати на запитання щодо достовірності інформації, пропонуючи перевірені факти та посилання на відповідні статті. Основні функції такі: «Перевірка фактів» – аналізує твердження та надає інформацію про їхню правдивість; «Автоматичні відповіді» – інтегрується з платформами, такими як Twitter або Facebook, для надання миттєвих відповідей.

FactCheck.org відноситься до некомерційних ресурсів, що спрямовані на перевірку достовірності заяв політиків, урядових діячів та медіа з метою зменшення дезінформації. Пропонуються такі функції: «Аналіз політичних заяв» – перевірка фактів у виступах, дебатах і рекламних кампаніях; «Спростування дезінформації» – викриття неправдивих новин і чуток; «Освітні ініціативи» – матеріали для навчання критичному мисленню.

ChatGPT є ботом генеративного штучного інтелекту, який здатний проводити інтерактивні бесіди, створювати текстовий контент і допомагати користувачам в різноманітних ситуаціях, використовуючи при цьому такі функціонали: «Автоматичні відповіді» – надання миттєвих текстових відповідей на запити; «Аналіз інформації» – допомагає в розумінні складних концепцій, пояснюючи їх доступною мовою; «Підтримка різних платформ» – інтегрується з численними додатками та сервісами, дозволяючи користувачам отримувати відповіді прямо в чатах, месенджерах або на вебсайтах; «Виконання завдань» – здатний виконувати завдання, такі як написання текстів, редагування, переклади та інші текстові маніпуляції.

До другої групи засобів виявлення інформаційного впливу на користувачів глобальної мережі Internet нами включено засоби пошуку оригінальних версій зображень, а, саме: Search by Image [10], TinEye [11], Google Image [12], Lenso.ai [13].

Search by Image є інструментом для пошуку за допомогою зображення, який дозволяє знаходити аналогічні або оригінальні зображення в Інтернеті через інтеграцію з пошуковими системами, такими як Google Images та Bing. Забезпечує такі функції як: «Пошук за зображенням», тобто можливість завантаження файлу зображення або вставка URL для знаходження схожих або ідентичний зображень; «Інтеграція з пошуковими системами» для підтримки популярних платформ з пошуку зображень; «Відкритий код» як код інструменту доступний на GitHub для модифікацій.

TinEye – забезпечує пошук за зображенням, який дозволяє знаходити оригінальні версії зображень та їх використання в Інтернеті за допомогою таких режимів: «Пошук за зображенням» для завантаження зображення або вставки URL для пошуку схожих або оригінальних зображень; «Зворотний пошук зображень» для пошуку джерел, де зображення було використано; «Пошук за історією» надає пошук посилання на першоджерело зображення.

Google Image – пошуковий інструмент – системи Google, для пошуку зображень та для знаходження зображень на основі текстових запитів або за допомогою зворотного пошуку за зображенням. Реалізує такі функції як «Пошук за текстовим запитом» для знаходження зображень за ключовими словами; «Пошук за зображенням» надає можливість завантажити зображення або вставити URL для пошуку схожих або оригінальних зображень; «Фільтри для результатів» надає опції для уточнення пошуку за розміром, типом, кольором та іншими характеристиками.

Lenso.ai використовує штучний інтелект для зворотного пошуку на основі введених користувачем запитів або пошуку аналогічних або ідентичних зображень за допомогою вхідного файлу. Дозволяє проводити: «Глибокий аналіз зображень» шляхом розпізнання обличчя, місць, об'єктів та навіть стилів зображень; «Відстеження джерел зображень» для визначення джерела, де вперше з'явилося зображення, «Фільтрацію результатів за доменом, ключовими словами та іншими параметрами.

Третю групу утворюють засоби автоматизованого виявлення маніпуляцій з медіа-контентом, до яких включимо: TrueMedia.org [45], AI Image Detector[15], Illuminarty [16], AI Detect Content [17], Wasitai [18]. Розглянемо окремі з них.

TrueMedia.org ефективно виявляє маніпуляції з медіа-контентом шляхом «Аналізу медіа-контенту», тобто перевірки зображень та відео на наявність змін або маніпуляцій та «Виявлення фальшивих матеріалів» – виявлення підробок, маніпуляцій з контентом та фактів фотошопу.

AI Image Detector дозволяє аналізувати зображення для визначення їх штучно згенероване походження (режим «Виявлення AI-генерованих зображень», а також завантажувати зображення і отримувати результат аналізу за кілька секунд (режим «Швидка перевірка»).

Окремо нами виділена четверта група засобів, яка об'єднала інструменти ручного виявлення маніпуляцій з медіа-контентом. Це 29PhotoForensics [19], FotoForensics [20], Imageforensic [21], Fake Image Detector [22].

29PhotoForensics є онлайн-інструментом для аналізу зображень, який допомагає виявляти маніпуляції. Він пропонує різні функції аналізу, включаючи перегляд метаданих і оцінку якості JPEG, що дозволяє визначати, чи було зображення змінено.

FotoForensics являє платформу для оцінки автентичності зображень, що включає аналіз метаданих та розширені алгоритми перевірки змін. Вона широко використовується для розслідувань у сфері дезінформації та перевірки фактів. Функціонал платформи об'єднує режими «Аналізу помилок (ELA)», «Перевірку метаданих EXIF», «Функцію порівняння з оригіналом».

Imageforensic надає засоби для експертного аналізу зображень, який зосереджений на автоматизованому виявленні підробок. Він забезпечує інтегровані сервіси для різних видів перевірок: «Розширений OCR» розпізнає текст у зображеннях, навіть якщо він був змінений; «Інструмент виявлення копіювання-вставки» визначає повторювані елементи, які могли бути використані для маніпуляцій; «Аналіз текстур і шумів» допомагає виявити неприродні елементи в цифровому зображенні.

Fake Image Detector створений для швидкої перевірки автентичності зображень, зокрема виявлення зображень, створених або змінених штучним інтелектом.

Засоби встановлення асоціації IP-адрес з країною або провайдером послуг утворили п'яту групу, до якої включили: IPlocation.net [23], That'sThem IP Address Lookup [24], IP FingerPrints [25], Who.is [26], 2IP [27], KeyCDN IP Location Finder [28], Geolocation.com [29], MaxMind's GeoIP Databases [30].

Iplocation.net, KeyCDN IP Location Finder, Geolocation.com, MaxMind's GeoIP Databases дозволяють визначати географічне місцезнаходження за IP-адресом, використовуючи при цьому різні джерела для точного визначення країни, міста, провайдера.

That'sThem IP Address Lookup дозволяє здійснювати зворотний пошук за IP-адресою, що дозволяє отримати інформацію про домени та вебсайти, пов'язані з конкретною IP-адресою.

IPFingerPrints дозволяє визначати характеристики конкретної IP-адреси через її аналіз та ідентифікацію пристроїв.

Who.is та 2IP забезпечують пошук розширеної інформації про домени та IP-адреси.

В останню, шосту групу інструментів нами включено засоби визначення довіри до адреси електронної пошти, які об'єднали Email-Checker.net [31], EmailHippo [32], MailboxValidato [33]r, That'sThem Reverse Email Lookup [34], Zero Bounce Free Email Verifier [35].

Email-Checker.net-дозволяє здійснювати перевірку існування електронної адреси – аналіз DNS-записів і SMTP-запитів.

EmailHippo надає розширені можливості перевірки електронних адресів (режим «Перевірка статусу електронної адреси») із додатковими даними про безпеку (режим «Оцінка ризиків») та активність облікових записів (режим «Інтеграція з системою управління взаємовідносинами з клієнтами (CRM)»).

MailboxValidator перевіряє наявність електронної пошти в реальному часі, допомагаючи уникнути проблем із доставкою листів (функції «Валідація електронної пошти», «Видалення дублікатів», «Інтеграція API»**).**

That'sThem Reverse Email Lookup дозволяє проводити зворотний пошук за електронною адресою для визначення додаткових даних про власника адреси.

Zero Bounce Free Email Verifier, MailMeteor Email Checker, CaptainVerify Email Checker забезпечують фільтрацію списків електронних адрес з метою підвищення рівня безпеки доставки електронних листів.

Проведемо оцінювання ефективності перерахованих вище засобів виявлення інформаційного впливу на учасників глобального інформаційного простору з використанням методу сценарного аналізу. Сценарні підходи допомагають, як правило, з'ясувати майбутні перспективи. Нами було розроблено сценарії практичної оцінки для кожної групи засобів інформаційного впливу. Кожний сценарій поєднував такі елементи як мета, вхід, кроки, очікувані результати, результати, подальша обробка. Результати оцінювання зводились в таблицю. Подібність сценаріїв дозволило навести в статті в якості прикладу лише один, наприклад, сценарій оцінки засобів першої групи, тобто оцінки засобів перевірки фактів у повідомленнях та новинах.

Сценарій оцінки засобів перевірки фактів у повідомленнях та новинах.

Мета**:** оцінити здатність інструментів перевірки фактів визначати правдивість статей новин за допомогою ключових слів, згенерованих штучним інтелектом.

Вхід: реальні (TRUE) та сфальсифіковані новини (FALSE):

1. Атака на Київську область: поліція показала фото з руйнуваннями [36] – TRUE;

2. No ‘red lines’ in Ukraine support, French foreign minister tells BBC [37] – TRUE;

3. Київ у шоці. Мешканці Покровська повідомили, що очікують російських визволителів [38] – FALSE;

4. Russian troops grind towards Pokrovsk, take second Ukrainian town in a week [39] – FALSE.

Кроки.

1. Вибір новини.

2. Створення набору релевантних ключових слів із тексту новини за допомогою шаблону діалогу для ChatGPT: «Привіт. Створи, будь-ласка, список з найбільш релевантних ключових слів для статті посилання на статтю»

3. Введення ключових слів в інструмент перевірки фактів.

4. Фіксація результату оцінки інструменту.

Очікувані результати. Інструменти перевірки фактів мають відповідати класифікаційній мітці, наданій у наборі даних (наприклад, якщо стаття позначена як «FALSE», то інструмент має ідентифікувати її як неправдиву або ненадійну, а якщо «TRUE», то – як правдиву).

Результати. Результати оцінювання кожного інструменту зводяться у таблицю.

Подальша обробка. Сортування інструментів за їхніми оцінками.

Група 1 − засоби перевірки фактів у повідомленнях та новинах. Таблиця оцінювання інструментів групи 1 за визначеною методикою доступна за посиланням https://tinyurl.com/3p8hf75a. Для обробки даних були застосовані Google – таблиці. Дані представлені на вкладці «Засоби перевірки фактів у повідомленнях і новинах», а результати проведених розрахунків оцінювання даних оформлені у вкладці «Результати». Як бачимо, такі інструменти як ChatGPT, FactCheck.org показали ефективність у перевірці достовірності тверджень з точністю до 75 % (табл. 1).

*Таблиця 1*

**Результати оцінки інструментів перевірки фактів у повідомленнях та новинах**

| Засіб | Відсоток відповідної (TRUE), невідповідної (FALSE) оцінки фактів та неможливість (UNDEFINED) надати оцінку | | |
|---|---|---|---|
| | TRUE | FALSE | UNDEFINED |
| ChatGPT | 75 % | 23 % | 3 % |
| FactCheck.org | 55 % | 30 % | 15 % |
| PolitiFact | 50 % | 20 % | 30 % |
| FactBot by Snopes | 33 % | 30 % | 38 % |

Отже, зазначені інструменти добре працюють із політичними заявами, новинами та онлайн-контентом. Проте в складних випадках, пов’язаних із емоційними тригерами чи багатошаровим контекстом, втрачають точність оцінки фактів.

Нами також розроблено сценарії для проведення практичного оцінювання засобів інших п’яти груп та одержані результати оцінки виявлення інформаційного впливу на користувачів мережі Інтернет. Проаналізуємо результати оцінювання.

Група 2 – засоби пошуку оригінальних версій зображень. Таблиця оцінювання інструментів за визначеною методикою доступна за посиланням https://tinyurl.com/yajn734w. Дані представлені на вкладці «Засоби перевірки фактів у повідомленнях і новинах», а результати проведених розрахунків оцінювання даних оформлені у вкладці «Результати». Отже, інструменти на кшталт Google Images, Lenso AI та GitHub Search by Image забезпечують високу точність (до 97 %) (табл. 2) у визначенні оригінальних зображень та виявленні їхніх копій.

*Таблиця 2*

**Результати оцінки засобів пошуку оригінальних версій зображень**

| Засіб | Відсоток відповідної (TRUE), невідповідної (FALSE) оцінки пошуку оригінальних версій зображень та неможливість (UNDEFINED) надати оцінку | | |
|---|---|---|---|
| | TRUE | FALSE | UNDEFINED |
| Search by Image | 97 % | 3 % | 0 % |
| Google Image | 93 % | 7 % | 0 % |
| Lenso.ai | 80 % | 20 % | 0 % |
| TinEye | 67 % | 33 % | 0 % |

Проте виявлено проблеми зі швидкістю обробки та точністю для зображень, які піддавалися серйозним модифікаціям або розмиттю метаданих.

Група 3 - засоби автоматизованого виявлення маніпуляцій з медіа-контентом. Таблиця оцінювання інструментів за визначеною методикою доступна за посиланням https://tinyurl.com/mwvvkyds. Результати оцінювання даних, наведені за вищезазначеним посиланням, показали, що інструменти, такі як Illuminarty та TrueMedia.org, демонструють надзвичайно високі результати (до 100 %) (табл. 3) у виявленні зображень, створених штучним інтелектом.

Водночас виклики виникають при аналізі гібридних зображень, створених шляхом поєднання реальних і згенерованих елементів. Це потребує додаткових досліджень і розробки покращених алгоритмів.

*Таблиця 3*

**Результати оцінки засобів автоматизованого виявлення маніпуляцій з медіа-контентом**

| Засіб | Відсоток відповідної (TRUE), невідповідної (FALSE) оцінки засобів для автоматизованого виявлення маніпуляцій з медіа-контентом та неможливість (UNDEFINED) надати оцінку | | |
|---|---|---|---|
| | TRUE | FALSE | UNDEFINED |
| Illuminarty | 100 % | 0 % | 0 % |
| TrueMedia.org | 97 % | 3 % | 0 % |
| AI Detect Content | 90 % | 10 % | 0 % |
| AI Image Detector | 80 % | 20 % | 0 % |
| Wasitai | 73 % | 27 % | 0 % |

Група 4 – засоби встановлення асоціації IP-адрес з країною або провайдером послуг. Таблиця оцінювання інструментів за визначеною методикою доступна за посиланням https://tinyurl.com/2m5x677d. Зібрані дані для даної групи та отримані результати оцінювання за вказаним посиланням засвідчили, що сервіси, такі як IPLocation, MaxMind та KeyCDN IP Location Finder, показали високу ефективність (до 100 %) (табл. 4) у визначенні геолокації IP-адрес.

*Таблиця 4*

**Результати оцінки засобів встановлення асоціації IP-адрес з країною або провайдером послуг**

| Засіб | Відсоток відповідної (TRUE), невідповідної (FALSE) оцінки засобів для вставлення асоціацій IP-адрес та неможливість (UNDEFINED) надати оцінку | | |
|---|---|---|---|
| | TRUE | FALSE | UNDEFINED |
| Iplocation.net | 100 % | 0 % | 0 % |
| MaxMind’s GeoIP Databases | 87 % | 0 % | 13 % |
| KeyCDN IP Location Finder | 83 % | 0 % | 17 % |
| IPFingerPrints | 80 % | 0 % | 20 % |
| Geolocation.com | 73 % | 0 % | 27 % |
| 2IP | 70 % | 0 % | 30 % |
| That’sThem IP Address Lookup | 63 % | 7 % | 30 % |
| Who.is | 10 % | 73 % | 17 % |

Група 5 – Засоби визначення довіри до адреси електронної пошти. Таблиця оцінювання інструментів за визначеною методикою доступна за посиланням https://tinyurl.com/ykemu6tf. Оцінювання даних даної групи дало можливість визначити, що такі інструменти як Email Checker, Mailbox Validator та ZeroBounce, досягли високого рівня точності (до 97 %) (табл. 5) у перевірці істинності електронних адрес.

Група 6 – Засоби ручного виявлення маніпуляцій з медіа-контентом. Таблиця оцінювання інструментів за визначеною методикою доступна за посиланням https://tinyurl.com/2p8er9v5. Проведене оцінювання дозволило встановити, що платформи, такі як 29PhotoForensics та FotoForensics, показали ефективність у виявленні змін у зображеннях із точністю до 83 % (табл. 6).

*Таблиця 5*

**Результати оцінки засобів визначення довіри до адреси електронної пошти**

| Засіб | Відсоток відповідної (TRUE), невідповідної (FALSE) оцінки засобів перевірки довіри до електронної пошти та неможливість (UNDEFINED) надати оцінку | | |
|---|---|---|---|
| | TRUE | FALSE | UNDEFINED |
| MailboxValidator | 97 % | 3 % | 0 % |
| CaptainVerify Email Checker | 95 % | 5 % | 0 % |
| Email-Checker.net | 93 % | 7 % | 0 % |
| EmailHippo | 93 % | 7 % | 0 % |
| That’sThem Reverse Email Lookup | 93 % | 7 % | 0 % |
| Zero Bounce Free Email Verifier | 93 % | 7 % | 0 % |
| MailMeteor Email Checker | 92 % | 8 % | 0 % |

Однак виклики виникають із тимчасовими адресами та новими доменами, для яких недостатньо історичних даних.

*Таблиця 6*

**Результати оцінки засобів ручного виявлення маніпуляцій з медіа-контентом**

| Засіб | Відсоток відповідної (TRUE), невідповідної (FALSE) оцінки засобів ручного виявлення маніпуляцій з медіа-контентом (UNDEFINED) надати оцінку | | |
|---|---|---|---|
| | TRUE | FALSE | UNDEFINED |
| Fake Image Detector | 83 % | 17 % | 0 % |
| FotoForensics | 83 % | 17 % | 0 % |
| Imageforensic | 80 % | 20 % | 0 % |
| 29PhotoForensics | 80 % | 20 % | 0 % |

Вони добре працюють із аналізом метаданих та перевіркою на помилки в JPEG. Проте інструменти стикаються зі складнощами при аналізі зображень з високоякісною ретушшю або зміненою текстурою, що ускладнює їхню ідентифікацію.

**Висновки та перспективи досліджень**

Отже, в рамках виконання завдань дослідження:

– проаналізовані сучасні методи та засоби виявлення інформаційного впливу, включаючи алгоритмічні рішення, системи штучного інтелекту та методи аналізу поведінкових моделей;

– розроблені сценарії практичної оцінки засобів виявлення інформаційного впливу, яка дозволяє комплексно оцінити їхню ефективність за визначеними критеріями;

– здійснено практичну оцінку визначених засобів, що дозволило виявити їхні сильні сторони та обмеження в контексті реальних сценаріїв інформаційного впливу.

Отримані результати можуть бути використані для вдосконалення існуючих методів і розробки нових інструментів з метою протидії інформаційним загрозам у глобальній мережі Інтернет.

**ЛІТЕРАТУРА**


[1] Number of internet and social media users worldwide as of October 2024 (in billions). Statista. URL: https://www.statista.com/statistics/617136/digital-population-worldwide/ (дата звернення: 03.01.2024).

[2] Дудатьєв А., Куперштейн Л., Войтович О. Інформаційне протиборство: моделі реалізації та оцінювання інформаційних операцій. *Кібербезпека: освіта, наука, техніка.* 2023. Т. 4, № 20. С. 72–80. https://doi.org/10.28925/2663-4023.2023.20.7280

[3] Литвин Н.А., Ярош А.О. Вплив дезінформації на національну безпеку України в умовах воєнного стану. *Юридичний науковий електронний журнал*. 2024, №2. С. 288–290. https://doi.org/ 10.32782/2524-0374/2024-2/69

[4] Гладка, О., Карпович, І., Тимракевич, А. Методика оцінювання впливу інформаційних загроз в умовах інформаційного протиборства. *Кібербезпека: освіта, наука, техніка*. 2024. Т. 4, № 24. С. 81–89. https://doi.org/10.28925/2663-4023. 2024.24.8189

[5] Алещенко В. Інформаційно-психологічна безпека особистості в умовах гібридної війни. *Вісник Київського національного університету імені Тараса Шевченка*. 2022, 1(49). С. 13–21. https://doi.org/10.17721/1728-2217.2022.49.13-21

[6] FactBot від Snopes. *Snopes:* веб-сайт. URL: https://www.snopes.com/factbot/ (дата звернення: 20.11.2024).

[7] PoliticoFact. The Poynter Institute: веб-сайт. URL: https://www.politifact.com/search/ (дата звернення: 20.11.2024).

[8] Factcheck.org. Проект Annenberg Public Policy Center: веб-сайт. URL: https://www.factcheck.org/ (дата звернення: 23.11.2024).

[9] Chatgtpt: веб-сайт. URL: https://chatgpt.com/ (дата звернення: 23.11.2024).

[10] GitHub: веб-сайт. URL: https://github.com/dessant/search-by-image (дата звернення: 24.11.2024).

[11] TinEye: веб-сайт. URL: https://tineye.com/ (дата звернення: 25.11.2024).

[12] Google зображення: веб-сайт. URL: https://images.google.com (дата звернення: 26.11.2024).

[13] Lenso.ai: веб-сайт. URL: https://lenso.ai/uk (дата звернення: 27.11.2024).

[14] TrueMedia.org: веб-сайт. URL: https://detect.truemedia.org/ (дата звернення: 27.11.2024).

[15] AI Image Detector: веб-сайт. URL: https://aiimagedetector.org/ (дата звернення: 29.11.2024).

[16] Illuminarty: веб сайт. URL: https://app.illuminarty.ai/#/ (дата звернення: 29.11.2024).

[17] AI Detect Content: веб-сайт. URL: https://aidetectcontent.com/ai-generated-image-detector/ (дата звернення: 30.11.2024).

[18] WasitAI: веб сайт. URL:https://wasitai.com/ (дата звернення: 30.11.2024).

[19] Forensically: веб-сайт. URL: https://29a.ch/photo-forensics/#forensic-magnifier (дата звернення: 02.12.2024).

[20] FotoForensics: веб-сайт. URL: https://fotoforensics.com/ (дата звернення: 02.12.2024).

[21] Ghiro: веб-сайт. URL: https://www.imageforensic.org/ (дата звернення: 03.12.2024).

[22] FID: веб-сайт. URL: https://www. Fakeimagedetector.com/ (дата звернення: 03.12.2024).

[23] IP Location: веб-сайт. URL: https://www.iplocation.net/ (дата звернення: 04.12.2024).

[24] That'sThem: веб-сайт. URL: https://thatsthem.com/reverse-ip-lookup (дата звернення: 04.12.2024).

[25] IP FingerPrints: веб-сайт. URL: https://www.ipfingerprints.com/ (дата звернення: 05.12.2024).

[26] WHO IS: веб-сайт. URL: https://who.is/ (дата звернення: 05.12.2024).

[27] 2IP: веб-сайт. URL: https://2ip.ua/ru/services/ information-service/whois (дата звернення: 06.12.2024).

[28] KeyCDN: веб-сайт. URL: https://tools.keycdn.com/geo (дата звернення: 06.12.2024).

[29] Geolocation: веб-сайт. URL: https://www.geolocation.com/ (дата звернення: 07.12.2024).

[30] MaxMind: веб-сайт. URL: https://www.maxmind.com/en/geoip-demo (дата звернення: 07.12.2024).

[31] Email-Checker: веб-сайт. URL:https://email-checker.net/check (дата звернення: 08.12.2024).

[32] EmailHippo: веб-сайт. URL: https://tools.emailhippo.com/ (дата звернення: 08.12.2024).

[33] MailboxValidator: веб-сайт. URL: https://www.mailboxvalidator.com/demo (дата звернення: 09.12.2024).

[34] That'sThem: веб-сайт. URL: https://thatsthem.com/reverse-email-lookup (дата звернення: 09.12.2024).

[35] ZeroBounce: веб-сайт. URL: https://www. zerobounce.net/free-email-verifier/ (дата звернення: 10.12.2024).
[36] Покровська В. Аналіз методів виявлення інформаційно-психологічного впливу в соціальних мережах. *Information Technology and Security*. 2020. 8(1) (14), 40–48. https://doi.org/10.20535/2411-1031.2020.8.1.218002
[37] Лаб'як І. Атака на Київську область: поліція показала фото з руйнуваннями. TCH.ua. URL: https://kyiv.tsn.ua/ataka-na-kiyivsku-oblast-policiya-pokazala-foto-z-ruynuvannyami-2710362.html (дата звернення: 19.12.2024).
[38] Київ у шоці. Мешканці Покровська повідомили, що очікують російських визволителів. *AiF*. URL: https://sensei-yoda.blogspot.com/2024/08/blog-post_134.html (дата звернення: 19.12.2024).
[39] Psaropoulos J. T. Russian troops grind towards Pokrovsk, take second Ukrainian town in a week. Aljazeera. URL: https://www.aljazeera.com/news/2024/10/9/russian-troops-grind-towards-pokrovsk-take-second-ukrainian-town-in-a-week (дата звернення: 19.12.2024).

**Носов В. В., Манжай О. В., Лучик С. Д., Лучик В. Є., Товстик В. О.**

## ПРАКТИЧНА ОЦІНКА ЗАСОБІВ ВИЯВЛЕННЯ ІНФОРМАЦІЙНОГО ВПЛИВУ НА КОРИСТУВАЧІВ ГЛОБАЛЬНОЇ МЕРЕЖІ INTERNET

*В статті представлені результати практичної оцінки ефективності доступних засобів виявлення інформаційного впливу на користувачів глобальної мережі. Зазначено, що внаслідок постійного збільшення потенціалу інтернет-середовища та надання користувачам практично необмежених можливостей в отриманні інформації виникає чимало загроз та ризиків для суспільства, які можуть мати довгострокові негативні наслідки: послаблення політичної стабільності держави, підрив довіри до інституцій або маніпулювання суспільною думкою.*

*Ідентифіковані та систематизовані сучасні методи та засоби виявлення інформаційного впливу на суспільство, включно алгоритмічні рішення, системи штучного інтелекту та методи аналізу поведінкових моделей. Виділено 6 груп засобів інформаційного впливу таких як перевірки фактів у повідомленнях та новинах, пошуку оригінальних версій зображень, автоматизованого виявлення маніпуляцій з медіа-контентом, інструменти ручного виявлення маніпуляцій з медіа-контентом, засоби встановлення асоціації IP-адрес з країною або провайдером послуг та визначення довіри до адреси електронної пошти.*

*Оцінювання ефективності визначених засобів виявлення інформаційного впливу на учасників глобального інформаційного простору проведено в статті з використанням методу сценарного аналізу. Нами розроблені сценарії практичної оцінки ефективності кожної групи засобів інформаційного впливу. Подібність сценаріїв дозволило навести в статті в якості прикладу лише один, зокрема, сценарій оцінки засобів перевірки фактів у повідомленнях та новинах.*

*Результати оцінювання за сценарним підходом показали, що більшість сучасних засобів, спрямованих на виявлення маніпуляцій з даними, демонструють прийнятну точність і функціональність у вирішенні завдань різної складності. Запропоновані авторами практичні підходи до оцінювання інформаційного впливу на користувачів можуть бути використані для вдосконалення існуючих сучасних методів і розробки нових інструментів з метою протидії інформаційним загрозам у глобальній мережі Інтернет.*

**Ключові слова:** мережа Internet; інформаційний вплив; ідентифікація; виявлення; оцінка; сценарій; ефективність; маніпуляції.

**Nosov V., Manzhai O., Luchyk S., Luchyk V., Tovstyk V.**

## PRACTICAL ASSESSMENT OF MEANS OF DETECTION OF INFORMATION IMPACT ON USERS OF THE GLOBAL INTERNET NETWORK

*The article presents the results of a practical assessment of the effectiveness of available means of detecting information impact on users of the global network. It is noted that as a result of the constant increase in the potential of the Internet environment and providing users with practically unlimited opportunities to obtain information, many threats and risks arise for society, which may have long-term negative consequences: weakening the political stability of the state, undermining trust in institutions or manipulating public opinion.*

*Modern methods and means of detecting information impact on society are identified and systematized, including algorithmic solutions, artificial intelligence systems and methods of analyzing behavioral models. 6 groups of information influence tools have been identified, such as fact-checking in messages and news, searching for original versions of images, automated detection of manipulations with media content, tools for manual detection of manipulations with media*

*content, tools for establishing an association of IP addresses with a country or service provider, and determining trust in an email address.*

*The article evaluates the effectiveness of the identified tools for detecting information influence on participants in the global information space using the scenario analysis method. We have developed scenarios for practical assessment of the effectiveness of each group of information influence tools. The similarity of the scenarios allowed us to cite only one example in the article, in particular, the scenario for evaluating tools for checking facts in messages and news.*

*The results of the scenario-based assessment showed that most modern tools aimed at detecting data manipulation demonstrate acceptable accuracy and functionality in solving tasks of varying complexity. The practical approaches proposed by the authors to assessing the information impact on users can be used to improve existing modern methods and develop new tools to counter information threats on the global Internet.*

**Keywords**: Internet, information impact, identification, detection, assessment, scenario, effectiveness, manipulation.

Стаття надійшла до редакції 11.01.2025 р.
Прийнято до друку 16.04.2025 р.