КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 3 (27), 2025
ISSN 2663 - 4023
DOI 10.28925/2663-4023.2025.27.720
УДК 004.056:519

Лавров Вадим Валерійович
аспірант кафедри захисту інформації
Вінницький національний технічний університет, Вінниця, Україна
vadim/lavrovv@gmail.com

Дудатьєв Андрій Веніамінович
к.т.н., доцент, доцент кафедри захисту інформації
Вінницький національний технічний університет, Вінниця, Україна
ORCID ID 0000-0002-7944-2404
dydatyev.av@gmail.com

ОГЛЯД ІСНУЮЧИХ МЕТОДІВ ОЦІНЮВАННЯ РИЗИКІВ ДЕЗІНФОРМАЦІЇ В УМОВАХ ГІБРИДНОЇ ВІЙНИ

Анотація. Стаття присвячена аналізу сучасних підходів до оцінювання ризиків дезінформації в умовах гібридної війни. Розглянуті основні технології автоматизованого виявлення фейкових новин, визначені основні прогалини підходів та запропоновані перспективи. Проведений огляд довів, що наявні моделі все ще недостатньо пристосовані до швидких змін у тактиках дезінформаторів, які використовують механізми «обману ШІ» й динамічно змінюють фейкові наративи. Серед ключових проблем, що уповільнюють розвиток ефективних систем протидії фейкам, названо відсутність локалізованих корпусів даних різними мовами, недостатньо визначені законодавчі норми й брак міждисциплінарного підходу, який би охоплював психологічні й соціальні аспекти сприйняття маніпулятивних повідомлень. Водночас дослідження підтвердили, що поєднання технологічних методів (машинне навчання, аналіз соцмереж, мультикритеріальна оцінка ризиків) із залученням експертного та користувацького досвіду дає змогу підвищити точність і швидкість ідентифікації фейкових новин, а також визначити пріоритетні заходи реагування. У висновках окреслено перспективи подальших розробок, зокрема мультимодальних систем детекції, здатних аналізувати відео- та аудіоконтент, а також рекомендації щодо інтеграції психологічних моделей, що вивчають когнітивні упередження аудиторії та її готовність сприймати пропаганду. Запропоновано низку підходів до підсилення міжнародного правового регулювання, яке могло б забезпечити ефективне стримування дезінформаційних атак без порушення базових прав людини й принципів свободи слова. Таким чином, стаття пропонує всебічний огляд сучасних досягнень і проблем у галузі оцінювання ризиків дезінформації, що може стати теоретичним і методичним підґрунтям для подальших прикладних і наукових розвідок у цій сфері. Результати можуть бути використані фахівцями в сфері інформаційної безпеки, дослідниками та органами влади для боротьби з деструктивним інформаційним впливом.

Ключові слова: гібридна війна; дезінформація; фейкові новини; машинне навчання; мультикритеріальний аналіз; deepfake; соціальні мережі; пропаганда; інформаційна безпека; когнітивні упередження; бот-мережі.

ВСТУП

Постановка проблеми. Дезінформація набула статусу одного з найдієвіших інструментів впливу в умовах гібридної війни, що визначається поєднанням військових, політичних, економічних та інформаційно-психологічних дій [1]. Зважаючи на надзвичайну швидкість поширення фейкових новин за допомогою соціальних мереж і різноманітних цифрових платформ, питання ефективного виявлення та оцінювання ризику дезінформації стає вкрай актуальним як у науковому, так і в практичному контексті. Одним із ключових завдань, яке постає перед дослідниками у цій сфері, є

розробка методів, здатних ідентифікувати фейкові повідомлення, прогнозувати їхній можливий вплив на аудиторію та надавати релевантну кількісну оцінку ризикам. Зважаючи на високі загрози для національної безпеки, політичної стабільності та суспільної довіри, важливість цього завдання важко переоцінити [2, с. 15]. У контексті гібридної війни, коли дезінформаційні операції здатні підірвати політичний устрій чи суспільний спокій, формування надійних моделей оцінювання ризику фейків є особливо актуальним.

Аналіз останніх досліджень і публікацій. У Питання дезінформації та способів її автоматизованого виявлення активно досліджувалося в роботах західних та вітчизняних авторів. Системні підходи до класифікації фейкових новин, включно з використанням машинного навчання, ґрунтовно розглянуто у працях [1] – [3]. Так, у [1] окреслено масштаб проблеми поширення фейків під час президентських виборів у США 2016 року та висвітлено соціальний вплив цих явищ. У [2] зроблено акцент на базових інструментах машинного навчання та формалізовано поняття «FakeNewsNet» як корпуса, що містить маркований фейковий і нефейковий контент. Додаткові аспекти застосування нейронних мереж, зокрема моделей на основі Convolutional Neural Networks та LSTM, висвітлено в [3]. Значний інтерес становлять роботи, присвячені ймовірнісним моделям і мультикритеріальному підходу. Наприклад, у [4], [5] детально розглянуто застосування Bayesian Networks, а також описано стратегії, які дезінформатори використовують, аби «обдурити» алгоритми детекції (adversarial attacks). У цих же публікаціях наголошується, що відсутність великих локалізованих датасетів суттєво ускладнює формування універсальних моделей. У вітчизняному контексті, особливо в світлі військової агресії та гібридних загроз, наявні дослідження [5], [6, с. 27] звертають увагу на специфіку національного інформаційного простору та культурно-лінгвістичні нюанси поширення фейків. Разом із тим, попри суттєвий прогрес, залишається низка аспектів, які не мають вичерпних рішень. Зокрема, існує потреба в подальшому вдосконаленні систем онлайн-моніторингу з використанням гібридних методів (комбінація машинного навчання, Bayesian Networks і мультикритеріального аналізу), аби якісно та кількісно оцінювати ризики у режимі реального часу [3], [5]. Крім того, все ще недостатньо розроблені моделі, які враховували б когнітивні упередження аудиторії та механізми поширення недостовірної інформації «зсередини» мережевих спільнот. Крім того, недостатньо дослідженим лишається інтегрований підхід, що дає змогу узагальнювати результати семантичних, ймовірнісних і мультикритеріальних методів в одну комплексну «панель» оцінювання ризику [5]. Відтак, саме ці не вирішені повною мірою питання і покликана розглянути дана стаття.

Метою статті є систематизація та аналіз існуючих підходів до оцінювання ризиків дезінформації з урахуванням реалій гібридної війни. У межах цієї мети передбачається:

- дослідити та узагальнити основні методи виявлення фейкових новин і пропагандистського контенту, включно з традиційними лексико-семантичними підходами, ймовірнісними моделями та мультикритеріальним аналізом;
- виокремити труднощі та прогалини в наявних дослідженнях, пов’язані з відсутністю локалізованих датасетів і недостатнім урахуванням когнітивних чинників;
- окреслити перспективні напрями, де комплексне поєднання різних інструментів (машинне навчання, Bayesian Networks, AHP/TOPSIS) дає змогу підвищити точність і оперативність оцінки ризиків, а також запропонувати рекомендації щодо подальших розробок із метою посилення захисту інформаційного простору.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Дезінформація сьогодні набула системного характеру, охоплюючи численні канали комунікації: телевізійні та друковані медіа, соціальні мережі, месенджери, онлайн-форуми й блоги. Суттєво посилилася загроза фейків і пропагандистських кампаній у контексті гібридних конфліктів, коли паралельно із військовими чи політичними діями активно використовується інформаційно-психологічний тиск. Високі темпи цифровізації й глобальна інтеграція соціальних мереж (Facebook, Twitter, Telegram, TikTok тощо) сприяють швидкому поширенню неправдивих чи викривлених повідомлень серед мільйонів користувачів. Одним із характерних прикладів аналізу масштабів цієї проблеми є дослідження [3], де було показано, що в період передвиборчих перегонів у США 2016 року фейкові новини могли досягати такої ж кількості переглядів, як і матеріали провідних ЗМІ. Наголошується, що одним із рушіїв масштабного поширення неправдивої інформації слугує так званий «вуглецевий слід» соціальних мереж, коли кожен користувач схильний ретранслювати матеріали, що підтверджують його вже сформовані переконання.

Важливе місце у вивченні сучасного стану дезінформації посідають напрацювання [4], у яких висвітлено комплекс проблем, пов’язаних із масованим використанням фейкового контенту, включно із застосуванням координованих бот-мереж. Автори підкреслюють, що нинішні стратегії дезінформаторів стають дедалі витонченішими: вони задіюють не лише звичні текстові фейки, а й синтезують аудіо, відео (технології deepfake), створюють цілі «екосистеми» фейкових акаунтів і груп, щоб формувати враження «масової підтримки» певних наративів. Окрему загрозу становлять труднощі у визначенні першоджерела дезінформації, що є надзвичайно актуальним у світлі гібридних конфліктів [2]. Багато ворожих операцій з поширення фейків розпочинається на анонімних майданчиках або спеціальних інтернет-форумах, де інформація «підживлюється» й підігрівається перед масовим запуском через великі соцмережі [8]. При цьому швидкість тиражування таких матеріалів у геометричній прогресії ускладнює їх оперативне спростування. Навіть за наявності фактчекінгу частина аудиторії встигає «увірувати» у фейкові наративи, і наступні спростування вже не мають такого резонансу. В українському та загалом східноєвропейському інформаційному просторі дезінформаційні кампанії часто спрямовані на підрив довіри до державних інституцій, виклик громадянських протестів чи посилення соціальної напруги. Деякі вітчизняні дослідження [5] звертають увагу на те, що такі інформаційні впливи можуть бути комплексними: разом із фейковими новинами поширюються дезорієнтуючі заяви «експертів», вирізані з контексту аудіо чи відеозаписи, а також цілеспрямовано організовані флешмоби в соціальних мережах. Таким чином, аналіз сучасного стану проблеми дезінформації засвідчує її складність і багатогранність. Масовість і висока швидкість поширення, технологічна витонченість (deepfake, бот-мережі, використання big data для таргетування), а також когнітивні упередження користувачів формують надзвичайно складне середовище, де традиційні механізми протидії (мануальний фактчекінг, заборона окремих ресурсів) виявляються недостатньо ефективними. Це, своєю чергою, викликає потребу в системному й науково обґрунтованому підході, що поєднує інформаційні, технологічні, соціологічні та правові методи.

МЕТОДИ АВТОМАТИЗОВАНОГО ВИЯВЛЕННЯ ФЕЙКОВИХ НОВИХ ТА ОЦІНЮВАННЯ РИЗИКІВ

Автоматизовані системи виявлення фейків та оцінювання пов'язаних із ними ризиків стали невіддільною складовою сучасних досліджень у сфері інформаційної безпеки. Розвиток цих методів можна умовно поділити на дві великі епохи: традиційного лексико-семантичного аналізу та більш сучасних алгоритмів машинного навчання (ML) і глибинного навчання (DL). У період традиційного лексико-семантичного підходу найчастіше застосовувалися словникові та статистичні методи: підрахунок частот ключових слів, аналіз синтаксичних патернів, виокремлення «сигнальних» фраз, що характерні для фейкових новин [3]. Такий підхід може допомогти помітити публікації із надмірно емоційним чи сенсаційним заголовком, однак його ефективність суттєво знижується, коли зловмисники уникають типових маркерів (наприклад, «шок», «терміново», «неповторно»). Окрім того, лексико-семантична модель, налаштована на одну мову чи культурний контекст, може не спрацювати в іншій мовній чи культурній площині через відмінності у фразеології та стилі.

На зміну традиційним методам прийшли алгоритми машинного навчання, які істотно розширили можливості детекції. Нейронні мережі (Convolutional Neural Networks, LSTM, Transformers) здатні аналізувати послідовність слів та враховувати більш тонкі семантичні й контекстуальні нюанси. Як приклад можна згадати роботу [5], де використано згорткові нейронні мережі для ідентифікації суперечливих або непослідовних тверджень у новинних публікаціях. Подальшого розвитку набули трансформер-архітектури (BERT, GPT, RoBERTa), які навчаються на величезних корпусах текстів і краще розпізнають тонкі маніпулятивні прийоми, що не завжди очевидні при поверхневому аналізі. Окрім власне аналізу тексту, сучасні моделі дедалі частіше використовують комплексні підходи, інтегруючи додаткові фактори:

- Метадані (хто є автором, як давно існує акаунт, чи помічено його в попередніх поширеннях фейків, наскільки активно поширюється контент).
- Соціальна мережа (структура зв'язків між користувачами, наявність груп ботів, швидкість «вірусного» розповсюдження).
- Фактчекінг від сторонніх організацій (визначення, чи встигли незалежні експерти назвати матеріал фейком і коли саме це сталося).

Суттєву роль в оцінюванні ризиків (risk assessment) відіграють моделі, що належать до мультикритеріального аналізу ймовірнісного характеру. Як приклад можна навести Байєсові мережі, застосовані в [4], де визначали ступінь фейковості контенту на підставі авторитетності джерела, емоційного забарвлення тексту та історії поширення. Ба більше, у деяких дослідженнях [2] пропонується поєднання методу AHP (Analytic Hierarchy Process) з машинним навчанням, щоби формувати інтегральну оцінку ризику. У таких системах ваги критеріїв (наприклад, достовірність джерела чи швидкість поширення) можуть бути визначені експертами, а результат класифікації, отриманий нейронною мережею, виступає одним із вхідних параметрів для мультикритеріальної моделі.

Ключовим викликом для всіх перелічених підходів є проблема масштабованості й динамічності. Зловмисники постійно вдосконалюють свої тактики, створюючи нові види контенту, тож моделі доводиться регулярно «донавчати» на актуальних даних [7]. Другою серйозною перешкодою є обмеженість локалізованих датасетів: більшість загальнодоступних корпусів (наприклад, FakeNewsNet) сфокусовані на англомовних

текстах і не завжди підходять для регіонів із власними лінгвістичними та культурними особливостями [5]. Як наслідок, навіть високоточні алгоритми можуть демонструвати знижену ефективність, коли переносяться на іншу мову чи інший культурний контекст.

Отже, сучасні методи автоматизованого виявлення фейкових новин і комплексного оцінювання ризиків поєднують у собі кілька шарів аналізу: лексико-семантичний, контекстуальний, соціальний та мультикритеріальний. Практика показує, що саме гібридний підхід (використання кількох алгоритмічних і методологічних «шарів» одночасно) дає найкращі результати в умовах гібридної війни, де швидкість і варіативність дезінформаційних атак постійно зростають. Комбінування машинного навчання, Байєсових мереж і мультикритеріальних моделей прийняття рішень дає змогу не лише детектувати фейкові повідомлення, а й надавати кількісну оцінку загроз, що дозволяє уповноваженим органам чи адміністраторам платформ пріоритизувати зусилля й реагувати оперативніше.

ПРОГАЛИНИ В ДОСЛІДЖЕННЯХ

Незважаючи на відчутний прогрес у галузі автоматизованого аналізу даних та машинного навчання, дослідники постійно вказують на низку проблем, що гальмують ефективну протидію дезінформації. Однією з найбільш суттєвих є обмеженість та неоднорідність доступних датасетів [2]. Велика частина існуючих корпусів для тренування й тестування алгоритмів детекції сфокусована на англомовному контенті, тоді як у контексті гібридної війни часто використовуються різні мови й регіональні діалекти [6]. При цьому автори робіт, зосереджених на локальних вибірках [5], наголошують, що фейкові наративи мають специфічні лексичні, синтаксичні та культурні риси, які ускладнюють розпізнавання на основі загальних моделей.

Іще одне «вузьке місце» пов’язане з визначенням і вимірюванням самого поняття «ризику дезінформації». Попри наявність метрик класифікаційної ефективності на кшталт Precision, Recall і F1-score [4], досить складно перевести їх у масштабі впливу на масову аудиторію чи безпекову ситуацію. Щоби оцінити реальний ризик, потрібно враховувати безліч чинників: охоплення фейків у соцмережах, швидкість поширення, стійкість аудиторії до контрнаративів, ступінь довіри до джерел, а також політичний чи соціальний контекст у конкретний момент часу. Окремо варто виділити обмеженість етичного та правового регулювання. Алгоритми машинного навчання здатні аналізувати велику кількість персональних даних, у тому числі поведінкові патерни користувачів [1], що потенційно загрожує порушенням приватності або може призвести до «превентивної цензури» задля блокування сумнівного контенту. Але в більшості країн законодавчі механізми або відсутні, або перебувають у стадії доопрацювання, і дослідники вимушені діяти в «сірій зоні» правового поля. Не менш важливою проблемою залишається здатність дезінформаторів вдосконалювати свої підходи [9]. Алгоритми «обману ШІ» (adversarial attacks) можуть цілеспрямовано змінювати стиль, формат або контент повідомлень таким чином, щоби обходити наявні фільтри. Зокрема, у деяких дослідженнях [5] продемонстровано, що вбудовані в текст певні «шуми» або маніпулятивне переставлення слів дають змогу істотно знизити точність навіть доволі потужних моделей глибинного навчання. Ці виклики підтверджують, що галузь далека від розв’язання проблеми дезінформації та вимагає подальшої розробки крос-дисциплінарних методів.

ПЕРСПЕКТИВИ РОЗВИТКУ

У контексті гібридної війни дослідники визначають кілька ключових напрямків, які можуть істотно посилити ефективність протидії дезінформації та сприяти розвитку сучасних методів оцінювання ризиків. Одним із першочергових завдань є створення й актуалізація локалізованих корпусів даних. Згадані вже збірки (FakeNewsNet, BuzzFeed, PolitiFact) зосереджені переважно на англомовних джерелах, тоді як необхідно систематично збирати «неангломовні» фейки з урахуванням регіональних особливостей політичного, історичного та культурного контексту [5]. Також все частіше порушується питання про мультимодальні алгоритми, які аналізують не лише текст, а й зображення, аудіо, відео. З огляду на збільшення кількості «deepfake»-роликів та маніпулятивних фото, гібридні техніки, що поєднують розпізнавання облич, голосу, текстової інформації та метаданих, стають критично важливими [2]. Іншим перспективним напрямком є інтеграція психологічних і соціальних моделей: врахування когнітивних спотворень, ефекту «ехо-камери», рівня довіри до джерел може підвищити точність прогнозування масштабів поширення та шкідливого впливу фейків.

Окремо варто звернути увагу на розробку комплексних систем, де багаторівневий аналіз (лексико-семантичний, ML, мультикритеріальний) поєднується з модулями фактчекінгу та онлайн-моніторингу соцмереж. Наприклад, у роботі [4] пропонується створення гібридної платформи, яка аналізує не лише контент, але й поведінкові патерни поширення (repost, like, retweet), а також орієнтується на показники мережевого впливу (кількість підписників, рівень взаємодії). Перспективною виглядає ідея впровадження адаптивних, «самонавчальних» систем, що у реальному часі оновлюють свої моделі, реагуючи на зміну методів дезінформаторів [10]. Ще одним важливим кроком є координація на міжнародному рівні та встановлення прозорих законодавчих рамок. Навіть найбільш досконалі алгоритми не дадуть бажаного результату, якщо не буде дієвого механізму відповідальності для платформ і користувачів, які систематично поширюють фейки. Тому формування міжнародних ініціатив та етичних стандартів (подібних до General Data Protection Regulation, але в контексті дезінформації) може суттєво змінити правила гри та стимулювати подальший розвиток наукових досліджень у цій сфері.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Результати огляду й аналізу існуючих методів оцінювання ризиків дезінформації в умовах гібридної війни підтверджують, що ця проблема має динамічний та багатогранний характер. Ефективна протидія фейковим новинам і пропагандистському контенту передбачає комплексний підхід, орієнтований не лише на автоматизовані системи моніторингу й аналізу текстової інформації, а й на врахування соціально-психологічних чинників і правових аспектів. Застосування алгоритмів машинного навчання, ймовірнісних моделей (зокрема Байєсових мереж) та мультикритеріальних методів (AHP, TOPSIS) суттєво поліпшує виявлення дезінформації та надає можливість оперативніше реагувати на появу нових фейкових наративів. Утім, недосконале правове та етичне регулювання породжує ризики використання ШІ як засобу цензури або порушення прав людини. Подальший розвиток досліджень має зосередитися на мультимодальних підходах, що дадуть змогу аналізувати відео, аудіо й зображення паралельно з текстами, а також інтегруватимуть когнітивні та психологічні моделі, аби

вивчати сприйнятливість різних аудиторій до маніпулятивного впливу. Водночас існує потреба в розширенні локалізованих датасетів різними мовами й удосконаленні законодавчих механізмів, покликаних забезпечити відповідальність платформ і користувачів у цифровому середовищі. Такий цілісний, адаптивний і відкритий для міждисциплінарної співпраці підхід сприятиме формуванню безпечного інформаційного простору й ефективному стримуванню дезінформаційних атак.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Wardle, C., Derakhshan, H., (2017). Information disorder: Toward an interdisciplinary framework for research and policy making.
2. Dudatyev, A., Kupershtein, L., & Voitovych, O. (2023). Information counterfeature: models of implementation and evaluation of information operations. *Electronic Professional Scientific Journal "Cybersecurity: Education, Science, Technique", 4(20),* 72–80. https://doi.org/10.28925/2663-4023.2023.20.7280
3. Filimowicz, M. (2023). *Information Disorder.* London: Routledge. https://doi.org/10.4324/9781003299936
4. Allcott, H., Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives, 31(2),* 211–236. https://doi.org/10.1257/jep.31.2.211
5. Dhiman, P. Kaur, A., Hamid, Y., & Ababneh, N. (2024). Fake News Detection Datasets: A Review and Research Opportunities. *International Journal of Computing and Digital Systems, 15(1),* 39–55. https://doi.org/10.12785/ijcds/160104
6. Korzh, O., & Korovai, B. (2023). Fake content: types, signs, ways of detection. *Education. Innovation. Practice, 11(7),* 37–42. https://doi.org/10.31110/2616-650X-vol11i7-005
7. Panda, B., Giordano, J. (1999). Defensive information warfare. *Communications of the ACM, 42(7),* 30–32. https://doi.org/10.1145/306549.306559
8. Davies, P. H. J. (2005). Intelligence, information technology, and information warfare. *Annual Review of Information Science and Technology, 36(1),* 312–352. https://doi.org/10.1002/aris.1440360108
9. Viitovych, T., & Syvakivskyi, Y. (2023). Hybrid Information Warfare in Modern Conditions. Social Communications: Tools, Technology and Practice. https://doi.org/10.36059/978-966-397-313-5-1
10. Boiko, V., & Burak, N. (2019). Information warfare as a threat to state security. *Protection of information in information and communication systems: Collection of abstracts of the III All-Ukrainian scientific and practical conference of young scientists, cadets and students,* 124–126.

КІБЕРБЕЗПЕКА: освіта, наука, техніка
CYBERSECURITY: EDUCATION, SCIENCE, TECHNIQUE
№ 3 (27), 2025
ISSN 2663 - 4023

Vadym Lavrov
Postgraduate Student of the Department of Information Security
Vinnytsia National Technical University, Vinnytsia, Ukraine
vadym@lavrovv@gmail.com

Andrii Dudatyev
PhD, Associate Professor, Associate Professor of the Department of Information Security
Vinnytsia National Technical University, Vinnytsia, Ukraine
ORCID ID: 0000-0002-7944-2404
dydatyev.av@gmail.com

REVIEW OF EXISTING METHODS FOR ASSESSING DISINFORMATION RISKS IN THE CONTEXT OF HYBRID WARFARE

Abstract. The article is devoted to analyzing modern approaches to assessing disinformation risks in the context of hybrid warfare. The primary technologies for automated fake news detection are reviewed, key gaps in the approaches are identified, and prospects for further research are proposed. The review confirms that existing models are still insufficiently adapted to the rapid changes in disinformation tactics, including "adversarial AI" techniques and dynamic shifts in fake narratives. Key issues hampering the development of effective systems for combating fake news include the lack of localized datasets in multiple languages, insufficiently defined legislative norms, and a lack of interdisciplinary approaches that integrate psychological and social aspects of perception of manipulative messages. At the same time, research has shown that combining technological methods (machine learning, social media analysis, multicriteria risk assessment) with expert and user input can significantly improve the accuracy and speed of identifying fake news while prioritizing response measures. The conclusions outline prospects for further developments, including multimodal detection systems capable of analyzing video and audio content, as well as recommendations for integrating psychological models to study audience cognitive biases and their readiness to accept propaganda. A range of approaches to strengthening international legal frameworks has been proposed, which could ensure effective containment of disinformation attacks without violating fundamental human rights and freedom of speech principles. Thus, the article provides a comprehensive review of current achievements and challenges in assessing disinformation risks, serving as a theoretical and methodological basis for further applied and scientific research in this field. The results can be used by information security specialists, researchers, and government authorities to counteract destructive informational influences.

Keywords: hybrid warfare; disinformation; fake news; machine learning; multicriteria analysis; deepfake; social media; propaganda; information security; cognitive biases; bot networks.

REFERENCES (TRANSLATED AND TRANSLITERATED)

1. Wardle, C., Derakhshan, H., (2017). Information disorder: Toward an interdisciplinary framework for research and policy making.
2. Dudatyev, A., Kupershtein, L., & Voitovych, O. (2023). Information counterfeature: models of implementation and evaluation of information operations. *Electronic Professional Scientific Journal "Cybersecurity: Education, Science, Technique", 4(20),* 72–80. https://doi.org/10.28925/2663-4023.2023.20.7280
3. Filimowicz, M. (2023). *Information Disorder.* London: Routledge. https://doi.org/10.4324/9781003299936
4. Allcott, H., Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives, 31(2),* 211–236. https://doi.org/10.1257/jep.31.2.211
5. Dhiman, P. Kaur, A., Hamid, Y., & Ababneh, N. (2024). Fake News Detection Datasets: A Review and Research Opportunities. *International Journal of Computing and Digital Systems, 15(1),* 39–55. https://doi.org/10.12785/ijcds/160104
6. Korzh, O., & Korovai, B. (2023). Fake content: types, signs, ways of detection. *Education. Innovation. Practice, 11(7),* 37–42. https://doi.org/10.31110/2616-650X-vol11i7-005

7. Panda, B., Giordano, J. (1999). Defensive information warfare. *Communications of the ACM, 42(7),* 30–32. https://doi.org/10.1145/306549.306559
8. Davies, P. H. J. (2005). Intelligence, information technology, and information warfare. *Annual Review of Information Science and Technology, 36(1),* 312–352. https://doi.org/10.1002/aris.1440360108
9. Viitovych, T., & Syvakivskyi, Y. (2023). Hybrid Information Warfare in Modern Conditions. Social Communications: Tools, Technology and Practice. https://doi.org/10.36059/978-966-397-313-5-1
10. Boiko, V., & Burak, N. (2019). Information warfare as a threat to state security. *Protection of information in information and communication systems: Collection of abstracts of the III All-Ukrainian scientific and practical conference of young scientists, cadets and students,* 124–126.