УДК 004.5:355

Д. С. Дженджеро, В. С. Наконечний, А. А. Побережний

МЕТОДОЛОГІЧНІ ПІДХОДИ ДО ВИЗНАЧЕННЯ ОЗНАК НЕПРАВДИВОЇ ІНФОРМАЦІЇ В СИСТЕМІ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ СИЛ ОБОРОНИ УКРАЇНИ

Обґрунтовано методологічні засади визначення ознак неправдивої інформації в системі інформаційної безпеки сил оборони України. Проаналізовано сучасні наукові підходи до ідентифікації дезінформації, зокрема когнітивний, контентно-семантичний, поведінковий, технічний та джерельний. Установлено, що у проведених дослідженнях зосереджено увагу на окремих аспектах проблеми, тоді як інтеграція цих підходів у єдину методологічну систему залишається недостатньо реалізованою. Запропоновано класифікацію п'яти груп ознак неправдивої інформації та методи їхнього кількісного оцінювання. Розроблено інтегральну модель оцінювання достовірності інформації, яка забезпечує перехід від якісного опису до формалізованої кількісної оцінки.

Ключові слова: *гібридна війна, дезінформація, фейкові повідомлення, достовірність інформації, інформаційна безпека, ознаки неправдивої інформації, модель оцінювання достовірності.*

Постановка проблеми. Гібридна війна є однією з найбільш складних і багатовимірних форм сучасного протистояння, у межах якої поєднуються військові, політичні, економічні, інформаційно-психологічні та кібернетичні засоби впливу, спрямовані на дестабілізацію системи державного управління та підрив національної стійкості.

В умовах довготривалої гібридної агресії проти України інформаційна безпека сил оборони набуває стратегічного значення. Противник активно застосовує технології дезінформації та психологічного впливу для спотворення оперативної обстановки, формування хибних наративів щодо бойових дій, втрат, логістики та морального стану військ. Такі інформаційні атаки спрямовані на дестабілізацію систем управління, зниження рівня довіри до командування та створення хаосу в процесах прийняття рішень. Тому визначення об'єктивних ознак неправдивої інформації має безпосереднє прикладне значення для системи інформаційної безпеки сил оборони, оскільки дає змогу підвищити надійність аналізу оперативних повідомлень і забезпечити інформаційну стійкість військових підрозділів.

Інформаційна складова гібридної війни набула системного характеру: вона охоплює цілеспрямоване використання дезінформації, пропаганди, симуляторів фейкових повідомлень, що впливають на масову свідомість, створюють викривлену реальність і формують сприятливий фон для політичних і військових рішень противника [1].

Дослідження сучасних конфліктів показують, що гібридна війна ґрунтується на «синергії загроз» – одночасному застосуванні військових, інформаційних та соціотехнічних методів впливу, де інформаційна компонента визначає швидкість і результативність операцій [2]. Саме це поєднання робить такі війни особливо небезпечними для демократичних держав, оскільки межа між традиційною агресією та інформаційно-когнітивним впливом стає практично непомітною. Отже, класичні засоби кіберзахисту, орієнтовані лише на технічні показники, виявляються недостатніми.

Паралельно у глобальному інформаційному просторі спостерігається експоненційне зростання масштабів і швидкості поширення неправдивої інформації. Аналіз сучасних дезінформаційних кампаній засвідчує, що вони розвиваються за закономірностями епідемічного поширення – фейки розповсюджуються швидше, ніж відбувається перевірка фактів, а багаторазове тиражування створює ілюзію достовірності [3]. Як наслідок виникає нова група семантичних ризиків, які не можуть бути виявлені виключно технічними засобами контролю кіберпростору.

Отже, актуальність дослідження зумовлена зростанням ролі дезінформації як ключового інструменту гібридної агресії, посиленням синергетичного характеру сучасних загроз та необхідністю розроблення цілісних методологічних підходів до визначення ознак неправдивої інформації, що забезпечать можливість її системного виявлення і протидії. Розроблення таких підходів є необхідною умовою підвищення ефективності захисту інформаційних ресурсів і формування інформаційної стійкості держави.

Аналіз останніх досліджень і публікацій. Питання визначення ознак неправдивої інформації в умовах гібридної війни дедалі частіше розглядається не лише як комунікаційна чи соціальна

проблема, а і як методологічне завдання ідентифікації та класифікації інформаційних впливів. Українські дослідники пропонують різні підходи до аналізу феномену дезінформації, проте більшість їхніх праць залишається у межах окремих дисциплін – права, соціальних наук або комунікацій – без створення єдиної інтегрованої системи критеріїв.

У статті [4] розглянуто когнітивно-комунікаційний підхід до виявлення неправдивих повідомлень, у межах якого визначаються поведінкові індикатори довіри і патерни поширення контенту в соціальних мережах. Акцентовано, що саме когнітивні реакції користувачів можуть виступати ознаками фейкових матеріалів, оскільки емоційна насиченість і повторюваність слів створюють ілюзію достовірності.

З правового погляду у дослідженні [5] розглянуто нормативно-оцінний підхід до визначення неправдивої інформації. Автори аналізують чинні законодавчі механізми та практики протидії дезінформації, показуючи, що сучасне регулювання здебільшого спрямоване на застосування санкцій і не містить чітко сформульованих методологічних критеріїв оцінки достовірності. Водночас наголошується на необхідності формування уніфікованих ознак неправдивості, які могли б бути інтегровані у систему інформаційної безпеки держави.

У міжнародному контексті дослідження методів протидії дезінформації відображене в аналітичному документі – звіті НАТО [6], де подано мультидисциплінарну модель оцінювання достовірності інформації. Вона передбачає взаємодію військових, наукових та громадських структур, що дає змогу поєднати технічні, семантичні й когнітивні методи аналізу повідомлень. Звіт визначає базові принципи верифікації даних, такі, як контекстна оцінка, перевірка джерела та ідентифікація повторюваних інформаційних шаблонів.

Розвиток методологічної бази чітко простежується у дослідженні [7], де дезінформація розглядається як частина когнітивно-наративного підходу до гібридної війни. Автори вводять концепт strategic narrative warfare, згідно з яким ознаки неправдивої інформації визначаються через аналіз смислових структур, суперечностей між фактами та наративною узгодженістю. Такий підхід дає змогу ідентифікувати інформаційні атаки на рівні змісту, а не лише фактології.

У праці [8] систематизовано контентно-семантичні та поведінкові підходи до виявлення неправдивої інформації в соціальних мережах. Автори узагальнюють лінгвістичні, семантичні та джерельні ознаки, а також патерни поширення контенту, що можуть свідчити про маніпулятивний або дезінформаційний характер повідомлень. Наголошується на доцільності комбінування цих ознак для підвищення точності автоматизованих методів детекції.

Когнітивно-поведінковий підхід до виявлення дезінформації подано у дослідженні [9], де увагу зосереджено на процесах сприйняття і довіри. Автори доводять, що ознаки неправдивої інформації можуть виявлятися не лише у змісті, а й у способі її споживання, зокрема, у швидкості поширення, гомогенності реакцій користувачів та зміні емоційного фону комунікації. Цей підхід розглядається як основа для побудови когнітивних моделей ризику дезінформації.

У сучасних працях фахівців сектору безпеки й оборони зростає увага до проблеми інформаційної достовірності у процесі бойового управління. Наголошується, що інформаційний складник операцій сил оборони має базуватися на верифікованих даних і механізмах перевірки достовірності у реальному часі. З огляду на досвід оборонних структур і зростання ролі інформаційної аналітики це зумовлює необхідність розроблення аналітичних моделей, здатних відокремлювати дезінформаційні повідомлення від оперативно достовірних, що безпосередньо впливає на ефективність планування, взаємодії та прийняття рішень у системі командування військами.

Узагальнюючи, можна зазначити, що сучасні дослідження зосереджуються на окремих підходах – правовому, когнітивному, контентно-семантичному чи технічному. Водночас їх інтеграція у єдину методологічну систему визначення ознак неправдивої інформації залишається недостатньо реалізованою. Саме потреба у такій уніфікації становить наукову нішу, у межах якої формується це дослідження.

Метою статті є розроблення концептуальних засад формування методологічних підходів до визначення ознак неправдивої інформації в умовах гібридної війни. Дослідження спрямоване на систематизацію застосовуваних наразі наукових підходів, виявлення їхніх сильних і слабких сторін та обґрунтування необхідності інтеграції когнітивних, семантичних, правових і технічних критеріїв у єдину методологічну систему.

Для досягнення поставленої мети передбачено вирішення таких основних завдань:

– проаналізувати сучасні підходи до класифікації та виявлення неправдивої інформації, що

застосовуються у сфері інформаційної безпеки;

– визначити ключові ознаки неправдивих повідомлень, релевантні до умов гібридного інформаційного впливу;

– обґрунтувати концептуальні принципи побудови цілісної методології ідентифікації неправдивої інформації, орієнтованої на підвищення ефективності систем захисту інформаційних ресурсів.

Виклад основного матеріалу. У межах сучасної гібридної війни визначення ознак неправдивої інформації потребує не лише описового, а і формалізованого підходу, який дає змогу здійснювати верифікацію повідомлень із заданою точністю. На основі проведеного аналізу було сформовано концептуальну модель, що поєднує класифікацію ознак, методи їхнього кількісного оцінювання та інтегральний показник достовірності інформації. Така побудова забезпечує можливість переходу від якісних характеристик до вимірюваних параметрів, що підвищує рівень об'єктивності виявлення неправдивих повідомлень.

Класифікація ознак неправдивої інформації. Систематизація ознак є базовим етапом побудови методології визначення неправдивості. У процесі дослідження виокремлено п'ять основних груп ознак, що охоплюють змістовий, структурний, поведінковий, технічний та джерельний аспекти інформації.

1. Контентні (семантичні) ознаки. Характеризують смислову структуру повідомлення, логічну узгодженість тверджень, рівень емоційної насиченості та використання маніпулятивних прийомів. До таких ознак належать: поляризована лексика, апеляції до страху або обурення, перебільшення, надмірні узагальнення.

2. Лінгвістичні ознаки. Визначаються статистичними характеристиками тексту: частотністю повторюваних слів, середньою довжиною речень, рівнем ентропії та відповідністю закону Ципфа.

Для природних текстів виконується умова

$$f(r) = \frac{C}{r^a}, \quad (1)$$

де $f(r)$ – частота слова з рангом r;

C – нормувальний коефіцієнт;

$a \approx 1$.

Значне відхилення від цього співвідношення вказує на штучність або маніпулятивність тексту.

3. Поведінкові ознаки. Відображують характер поширення повідомлення у мережі. Для неправдивої інформації типовими є різке зростання кількості репостів у короткий проміжок часу, синхронність появи однакових текстів на різних ресурсах, брак тематичної різноманітності у коментарях.

4. Технічні ознаки. Охоплюють параметри структури повідомлення і властивості джерела поширення (метадані, часові мітки, IP-адреси, взаємозв'язки між акаунтами). Вони формують основу для побудови мережевих графів інформаційних потоків.

5. Джерельні ознаки. Характеризують ступінь достовірності каналу походження інформації, історію його активності та співвідношення між фактичними й емоційними повідомленнями.

Методи визначення ознак неправдивої інформації. Для кожної групи ознак застосовано відповідний метод кількісного оцінювання, що дає змогу перевести якісні спостереження у вимірювані параметри.

Ентропійний аналіз. Для оцінювання структурної складності тексту використано ентропію Шеннона, яка визначає ступінь невпорядкованості мовних елементів:

$$H = -\sum_{i=1}^{n} p_i log_2 p_i, \quad (2)$$

де p_i – імовірність появи i-го слова або символу.

Природні тексти мають оптимальний рівень ентропії; надмірно висока або низька ентропія сигналізує про неприродність побудови повідомлення (автоматична генерація, навмисна спрощеність, шаблонність).

Частотний аналіз і закон Ципфа. Якщо для тексту не виконується закон Ципфа, то це свідчить про його штучність або спотворення лексичної рівноваги. Порівняння емпіричного розподілу частот із теоретичним дає змогу виявити інформаційні аномалії, що характерні для фейкових повідомлень.

Вагова оцінка ключових слів (TF-IDF). Для кількісної оцінки семантичної значущості термінів використовується формула

$$w_{i,j} = tf_{i,j} \times \log \frac{N}{df_i}, \quad (3)$$

де $tf_{i,j}$ – частота слова i у документі j;

df_i – кількість документів, у яких трапляється слово i;

N – загальна кількість документів.

Отримані ваги використовуються для визначення тематичних відхилень та оцінювання наративної схожості між джерелами.

Визначення подібності контенту. Для аналізу схожості повідомлень у межах однієї інформаційної кампанії застосовується метрика косинусної подібності:

$$\cos(\theta) = \frac{A \cdot B}{||A||\,||B||}, \quad (4)$$

де A і B – вектори ознак двох текстів.

Велике значення $\cos(\theta)$ (наближене до 1) означає значну схожість між повідомленнями, що вказує на координоване поширення контенту з єдиного джерела.

Поведінкове моделювання. Поведінкові індикатори визначаються за допомогою аналізу часових рядів активності користувачів.

Для оцінювання швидкості поширення повідомлення використовується коефіцієнт темпу реплікації

$$K_r = \frac{N_t - N_{t-1}}{\Delta t}, \quad (5)$$

де N_t – кількість публікацій або реакцій у момент часу t;

Δt – часовий інтервал спостереження.

Різке зростання K_r у короткий період указує на штучне стимулювання поширення контенту.

Інтегральна модель оцінювання достовірності інформації. Для узагальнення результатів аналізу запропоновано інтегральну модель визначення достовірності повідомлення, що базується на зваженому підсумовуванні часткових коефіцієнтів для кожної групи ознак:

$$D = \sum_{i=1}^{n} w_i k_i, \quad (6)$$

де D – інтегральний показник достовірності;

w_i – ваговий коефіцієнт значущості i-ї ознаки;

k_i – нормалізоване значення достовірності за цією ознакою.

Значення w_i визначаються експертно або на основі статистичних вибірок.

Кожна з п’яти груп ознак оцінюється у шкалі [0;1], де 1 відповідає максимально достовірному повідомленню, а 0 – типовій дезінформації. Такий підхід дає змогу кількісно порівнювати повідомлення за ступенем правдивості і виявляти потенційно небезпечні інформаційні вкиди.

Інтерпретація результатів здійснюється за шкалою:

– $D > 0{,}75$ – інформація достовірна;

– $0{,}45 < D \leq 0{,}75$ – інформація потенційно маніпулятивна і потребує перевірки;

– $D \leq 0{,}45$ – інформація неправдива або з високим рівнем викривлення.

Отриманий показник може бути використаний для автоматизованого моніторингу інформаційного простору, а також для формування рішень у системах кіберзахисту й аналітичних платформах інформаційної безпеки.

Отже, розроблена методологія уможливлює комплексне виявлення неправдивої інформації в умовах гібридної війни шляхом поєднання контентного, лінгвістичного, поведінкового та технічного аналізу. Її застосування сприяє підвищенню ефективності захисту інформаційних ресурсів, зміцненню інформаційної стійкості держави та створює підґрунтя для подальшої автоматизації процесів ідентифікації дезінформації.

Висновки

У результаті проведеного дослідження сформовано концептуальні засади методології визначення ознак неправдивої інформації в умовах гібридної війни. На основі аналізу наукових праць і практик ідентифікації дезінформації виявлено, що застосовувані наразі підходи здебільшого зосереджені на окремих аспектах – когнітивних, семантичних, правових або технічних – і не забезпечують

комплексної оцінки достовірності повідомлень. Це підтвердило необхідність розроблення інтегрованої системи критеріїв, яка б поєднувала різні рівні аналізу.

Запропонована методологія базується на системному підході і передбачає виокремлення п’яти груп ознак: когнітивно-семантичних, лінгвістичних, поведінкових, технічних та джерельних. Для кожної групи визначено набір параметрів, що дають змогу кількісно оцінювати рівень достовірності, а також розроблено інтегральну модель, яка забезпечує обчислення загального коефіцієнта правдивості повідомлення. Це уможливлює перехід від описового аналізу до формалізованої оцінки інформаційних впливів.

Практична значущість отриманих результатів полягає у створенні науково-методологічної основи для подальшого проєктування алгоритмів і систем автоматизованого виявлення дезінформації. Запропонована модель може бути використана для підвищення рівня інформаційної безпеки сил оборони України. Інтеграція розробленої моделі до систем інформаційного забезпечення та аналітичних підрозділів дасть змогу виявляти деструктивні інформаційні впливи на ранніх етапах, зменшувати ризик прийняття рішень на підставі викривлених даних та підвищувати інформаційну стійкість підрозділів у ході службово-бойової діяльності.

Подальші дослідження доцільно спрямувати на уточнення вагових коефіцієнтів груп ознак із використанням методів машинного навчання, розширення моделі на мультимедійний контент і дослідження динаміки когнітивних реакцій аудиторії як одного з ключових чинників поширення дезінформації.

Перелік джерел посилання

1. Варга Т. М. Дезінформація та пропаганда як інструменти ведення гібридної війни росії проти України. Київ : НУОУ, 2024. 28 с.
2. Vuković J., Matika D., Barić S. Hybrid Warfare Challenges. Zagreb : Croatian Defence Academy, 2022. 24 p.
3. Patel S. S., Maloney E. J., Omer S. B. The Landscape of Disinformation on Health Crisis Communication During the COVID-19 Pandemic in Ukraine. *Journal of Global Health.* 2020. Vol. 10 (2). P. 1–8. DOI: https://doi.org/10.7189/jogh.10.020310.
4. Ратушна І. Когнітивно-комунікаційні механізми сприйняття дезінформації в цифровому середовищі. *Інформаційне суспільство.* 2024. № 2 (16). С. 45–53.
5. Смотрич Д., Іванов Н. Правові аспекти боротьби з дезінформацією в Європейському Союзі: уроки для України. *Вісник Національного університету «Львівська політехніка». Юридичні науки.* 2023. № 4 (40). С. 155–161. DOI : https://doi.org/10.23939/law2023.40.155.
6. NATO StratCom COE. Disinformation Resilience Index: A Multidisciplinary Model of Credibility Assessment. Riga : NATO Strategic Communications Centre of Excellence, 2024. 60 p.
7. Bachmann S. D., Gunneriusson H., Giegerich B. Hybrid Warfare and Disinformation: A Ukraine War Perspective. *Defence Strategic Studies Journal.* 2023. Vol. 12 (1). P. 34–51. DOI: https://doi.org/10.2478/dssj-2023-0003.
8. Shen, H., Li, T., & Zhang, Y. Fake news detection on social networks: A survey. *Applied Sciences. (MDPI)*. 2023. Vol. 13 (21). DOI : https://doi.org/10.3390/app132111877.
9. Danyk Y., Briggs C. Modern Cognitive Operations and Hybrid Warfare. *Security and Defence Quarterly.* 2023. Vol. 42 (2). P. 55–68. DOI: https://doi.org/10.35467/sdq/168741.

Стаття надійшла до редакції 17.11.2025 р.

UDC 004.5:355

D. Dzhendzhero, V. Nakonechnyi, A. Poberezhnyi

METHODOLOGICAL APPROACHES TO IDENTIFYING THE FEATURES OF FALSE INFORMATION WITHIN THE INFORMATION SECURITY SYSTEM OF THE DEFENSE FORCES OF UKRAINE

The article substantiates the conceptual and methodological foundations for identifying the

characteristics of false information within the information security system of the Defense Forces of Ukraine, emphasizing the growing importance of information security and cognitive resilience in contemporary conflicts. The study systematizes and critically analyzes modern approaches to the detection of disinformation, integrating cognitive, content-semantic, behavioral, technical, and source-based dimensions. It is established that the majority of existing studies address only partial aspects of this phenomenon, lacking a unified methodological framework capable of combining qualitative and quantitative parameters of credibility assessment.

The proposed approach introduces a comprehensive classification of five groups of false information indicators and defines specific quantitative metrics for their evaluation. These include entropy and frequency analysis of textual data, Zipf's law for lexical distribution regularities, TF-IDF weighting for semantic relevance, cosine similarity for intertextual comparison, and the replication rate coefficient to model the dynamics of message dissemination. The integration of these components forms an integral credibility model that enables the transition from descriptive assessments to formalized, measurable evaluation of information veracity.

The developed methodological framework allows for a multi-layered analysis of information content and dissemination patterns, providing a more objective basis for detecting manipulative or disinformative narratives within hybrid influence operations. The results obtained contribute to the advancement of scientific understanding of disinformation mechanisms and create a foundation for the development of automated analytical systems capable of identifying hybrid information threats in real time.

The study's practical significance lies in its potential application for the design of intelligent information monitoring systems and early warning mechanisms against disinformation campaigns. Future research directions include the refinement of weighting coefficients for credibility indicators through machine learning, the extension of the model to multimedia content, and the exploration of cognitive response dynamics as a critical factor in the spread of false information.

Keywords: *hybrid warfare, disinformation, fake news, information credibility, information security, signs of false information, credibility assessment model.*

Дженджеро Дмитро Сергійович – аспірант кафедри кібербезпеки та захисту інформації, Київський національний університет імені Тараса Шевченка
https://orcid.org/0009-0007-9999-850X

Наконечний Володимир Сергійович – доктор технічних наук, професор, професор кафедри кібербезпеки та захисту інформації, Київський національний університет імені Тараса Шевченка
https://orcid.org/0000-0002-0247-5400

Побережний Андрій Анатолійович – науковий співробітник науково-дослідної лабораторії службово-бойового застосування НГУ, Національна академія Національної гвардії України
https://orcid.org/0000-0002-8984-6912