МВС України
Харківський національний університет внутрішніх справ
Департамент кіберполіції Національної поліції України
Львівський державний університет безпеки життєдіяльності

Практична оцінка засобів виявлення дезінформації

Науково-методичні рекомендації

Кам’янець-Подільський 2025

УДК 343.98:[343.54:053.2/6:004](477)
ББК 67.9 (4УКР) 623.13

Рекомендовано до друку Вченою радою
Харківського національного університету внутрішніх справ
(протокол № 4 від 26 листопада 2025 р.)

ЗАТВЕРДЖЕНО Науково-методичною радою Харківського національного університету внутрішніх справ від 19.11.2025 р. Протокол № 10	**СХВАЛЕНО** Вченою радою ННІ №4 Харківського національного університету внутрішніх справ від 18.11.2025 р. Протокол № 11
ПОГОДЖЕНО Секцією Науково-методичної ради Харківського національного університету внутрішніх справ з технічних дисциплін від 18.11.2025 р. Протокол № 10	**РОЗГЛЯНУТО** на засіданні кафедри протидії кіберзлочинності ННІ № 4 Харківського національного університету внутрішніх справ від 30.10.2025 р. Протокол № 23

Рецензенти:
О. М. Тарасов – заступник директора з розвитку ДП «Інфотех»;
Ю. В. Гнусов – завідувач кафедри кібербезпеки та DATA-технологій навчально-наукового інституту № 5 Харківського національного університету внутрішніх справ, кандидат технічних наук, доцент;

Практична оцінка засобів виявлення дезінформації : науково-методичні рекомендації / О.В. Манжай, В.В. Носов, С.В. Самойлов, Р.Л. Ткачук, В.О. Товстик. Х. : ХНУВС, 2025. 41 с.

Розробники: Манжай О.В., к.ю.н., професор; Носов В.В., к.т.н., доцент; Самойлов С.В., к.ю.н.; Ткачук Р.Л., д.т.н., професор; Товстик В.О.

КОЛЕКТИВ АВТОРІВ

Манжай Олександр Володимирович, завідувач кафедри протидії кіберзлочинності ННІ № 4 Харківського національного університету внутрішніх справ, кандидат юридичних наук, професор;
Носов Віталій Вікторович, професор кафедри протидії кіберзлочинності ННІ № 4 Харківського національного університету внутрішніх справ, кандидат технічних наук, професор;
Самойлов Станіслав Вадимович, начальник 3-го управління (інформаційних технологій та програмування) Департаменту кіберполіції Національної поліції України, кандидат юридичних наук;
Ткачук Ростислав Львович, професор кафедри управління інформаційною безпекою навчально-наукового інституту цивільного захисту Львівського державного університету безпеки життєдіяльності, доктор технічних наук, професор
Товстик Вадим Олександрович, курсант 4 курсу ННІ № 4 Харківського національного університету внутрішніх справ.

ЗМІСТ

ПЕРЕДМОВА

Стрімкий розвиток цифрових технологій та розширення можливостей глобальної мережі Інтернет суттєво змінили механізми поширення, сприйняття й використання інформації в сучасному суспільстві. У цифровому середовищі інформаційні потоки формуються та циркулюють у режимі реального часу, впливаючи на суспільну думку, індивідуальну поведінку користувачів та політичні процеси на національному й міжнародному рівнях. У таких умовах питання інформаційного впливу та, зокрема, дезінформації набуває не лише теоретичного, а й виразного практичного значення, оскільки від ефективності виявлення та нейтралізації деструктивного контенту безпосередньо залежить рівень інформаційної безпеки держави та стійкість суспільства до маніпулятивних загроз.

У сучасному інформаційному просторі дезінформація дедалі частіше використовується як інструмент гібридного впливу, спрямованого на дестабілізацію внутрішньої ситуації, підрив довіри до демократичних інститутів, формування хибних уявлень про події, політичні рішення чи дії уряду. Відкрита природа соціальних мереж, доступність механізмів автоматизації поширення контенту, використання технологій штучного інтелекту для створення глибоких фейків та маніпулятивних повідомлень значно ускладнюють оперативне виявлення недостовірної інформації. Водночас саме цифрове середовище створює умови для розроблення нових підходів та інструментів, здатних забезпечувати більш точне та швидке виявлення дезінформаційних матеріалів.

Важливість наукової оцінки сучасних засобів виявлення дезінформації обумовлена кількома взаємопов’язаними факторами. По-перше, обсяги інформації, яку споживають користувачі Інтернету, зростають у геометричній прогресії, що робить традиційні методи аналізу малоефективними. По-друге, структура інформаційного впливу стає дедалі складнішою, оскільки маніпулятивний контент дедалі частіше маскується під авторитетні джерела або поширюється через мережі ботів і тролів. По-третє, у суспільстві спостерігається зниження рівня критичного мислення, що підвищує вразливість користувачів до деструктивної інформації та її психологічних ефектів. У цих умовах постає

нагальна потреба в системному аналізі як теоретичних засад інформаційного впливу, так і практичних можливостей сучасних засобів протидії дезінформації.

Дослідження інформаційного впливу в глобальній мережі охоплює кілька взаємопов'язаних площин. Психологічні механізми визначають, як саме користувачі сприймають інформацію, на які когнітивні упередження реагують та які емоційні тригери використовують дезінформаційні кампанії. Соціальний вимір включає формування групових норм, поведінкові зміни, феномени інформаційних бульбашок і поляризації. Політичні наслідки інформаційних впливів можуть проявлятися у трансформації суспільної довіри, підриві авторитету державних інституцій, маніпуляції виборчими процесами та переформатуванні зовнішньополітичних орієнтирів. Усі зазначені аспекти формують комплексну проблему, яка потребує практичної оцінки ефективних засобів її нейтралізації.

Окремої уваги потребує аналіз сучасних технічних засобів виявлення дезінформації. Серед найбільш поширених слід виділити системи на основі машинного навчання, алгоритми обробки природної мови (NLP), методи виявлення аномалій у поширенні контенту, інструменти фактчекінгу та автоматизованого моніторингу соціальних мереж. Важливим є також дослідження їхніх обмежень: проблематичність інтерпретації алгоритмів, ризики хибних спрацювань, залежність від якості навчальних вибірок та складність адаптації систем до нових форм інформаційних загроз.

Таким чином, актуальність дослідження практичної оцінки засобів виявлення дезінформації визначається як загрозливим масштабом маніпулятивних впливів, так і потребою формування цілісної системи захисту суспільства від інформаційних загроз. Отже, у межах матеріалів науково-методичних рекомендацій основний акцент зосереджено на визначенні теоретичних засад дослідження, уточненні ключових понять та формуванні аналітичної рамки, у якій аналізуватимуться психологічні, соціальні й політичні наслідки інформаційного впливу. Особлива увага приділяється можливостям сучасних засобів виявлення дезінформації та окресленню найбільш ймовірних сценаріїв їх практичної оцінки.

1. ЗАГАЛЬНА ХАРАКТЕРИСТИКА ІНФОРМАЦІЙНОГО ВПЛИВУ НА КОРИСТУВАЧІВ ГЛОБАЛЬНОЇ МЕРЕЖІ

Інформаційні ресурси глобальної комп'ютерної мережі Інтернет стають домінантними у формуванні інформаційного простору. Великі масиви контенту, миттєва комунікація, соціальні платформи сприяють прискореному обігу даних. При цьому користувачі молодшої вікової групи й із високим рівнем освіти мають більшу здатність критично оцінювати інформацію. Дослідження показують [1], що низька цифрова грамотність супроводжується підвищеною вразливістю до маніпуляцій.

Станом на лютий 2025 [2] року близько 5,56 млрд осіб використовують інтернет — це приблизно 67,9% світового населення. У той самий період соціальними мережами користуються приблизно 5,24 млрд осіб або 63,9 % світової популяції (рис. 1.1.).

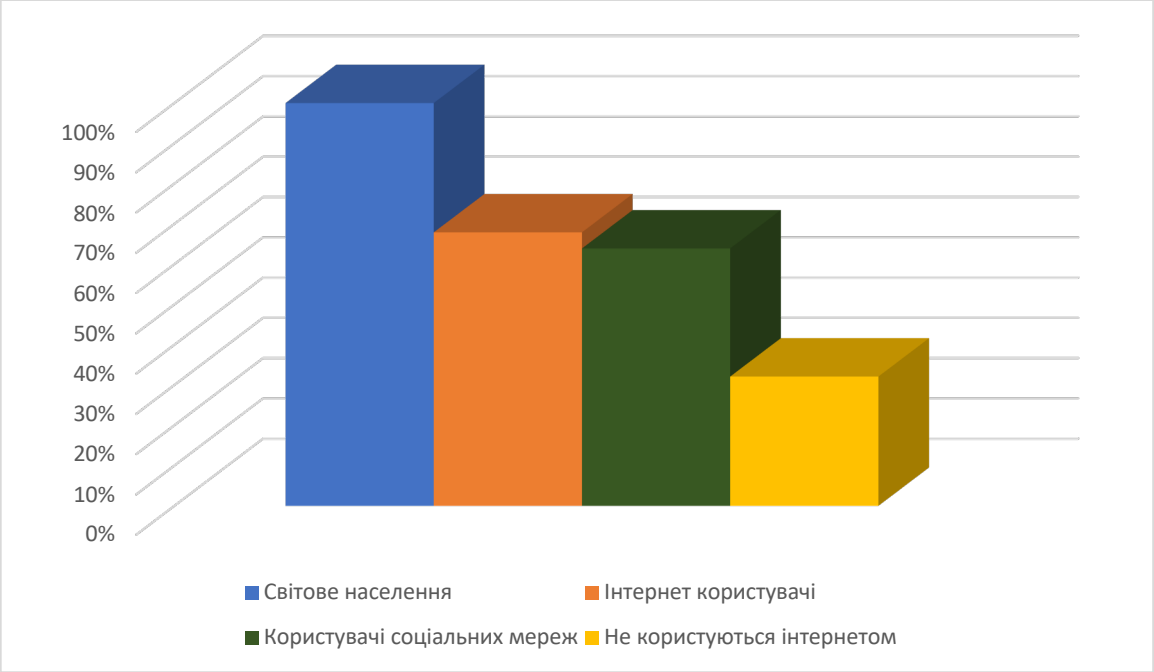

Рис. 1.1. Співвідношення світового населення до інтернет-користувачів та користувачів соціальних мереж за даними [2]

Молодь (15–24 років) демонструє найвищий рівень проникнення інтернету — в Європі [2] цей показник досягає 98%, тоді як у світі середній показник для цієї вікової групи становить 79% (рис. 1.2.).

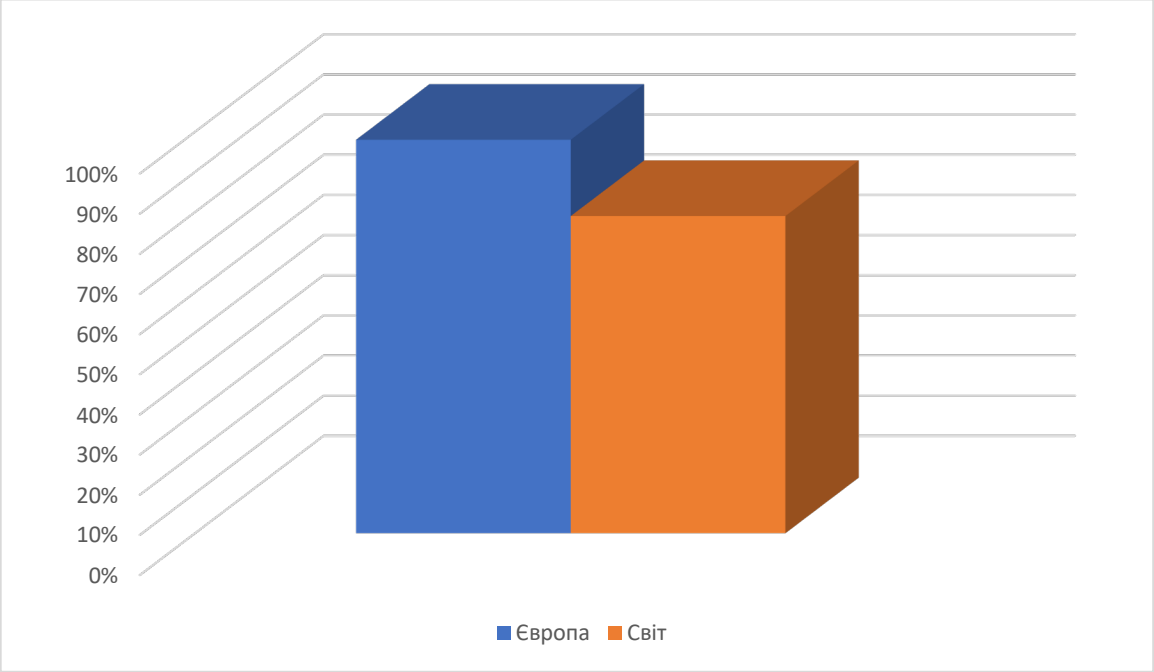

Рис. 1.2. Ступень проникнення інтернету серед молоді (15–24 років) за даними [2]

Одночасно у країнах із високим доходом інтернет-доступ мають приблизно 93% населення, тоді як в країнах із низьким доходом — близько 27% (рис 1.3.) [2].

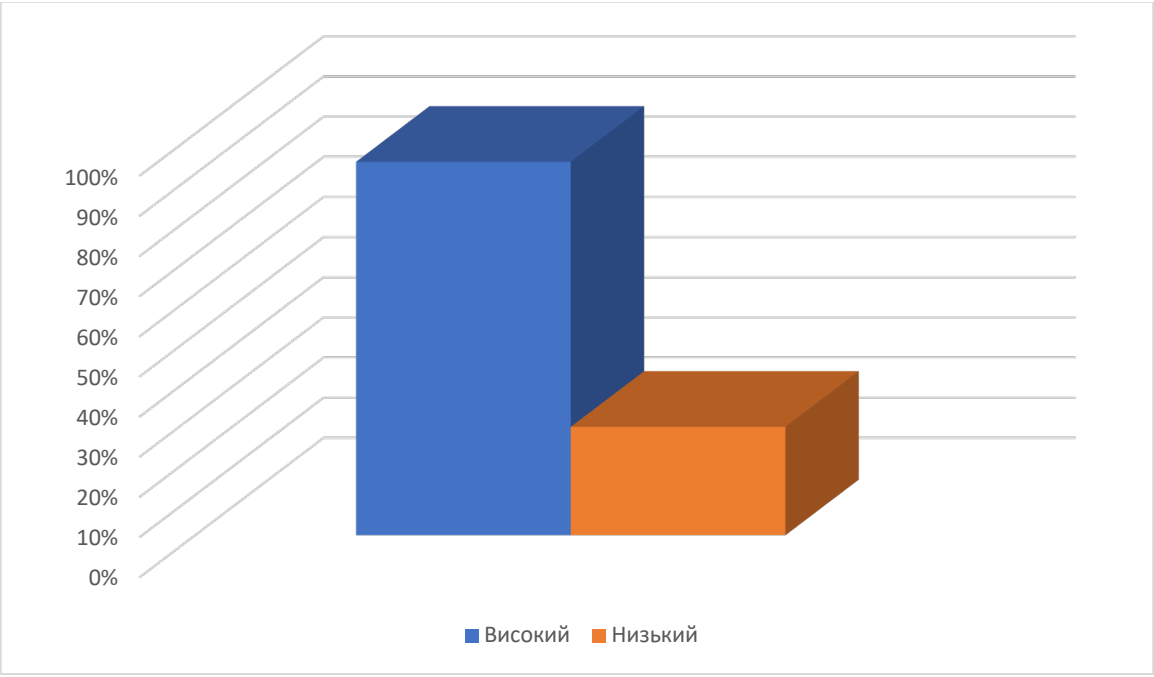

Рис. 1.3. Статистика доступності до інтернету за рівнем доходу за даними [2]

В Україні, за дослідженням Центру Разумкова [3] 72% громадян отримують

новини через соціальні мережі, з яких 62% — через Telegram. Щодня новини читають або переглядають 65% опитаних. Найбільший інтерес викликають теми: війна з РФ – 75 %, внутрішня політика – 48%, міжнародні події – 43,5% (рис. 1.4.).

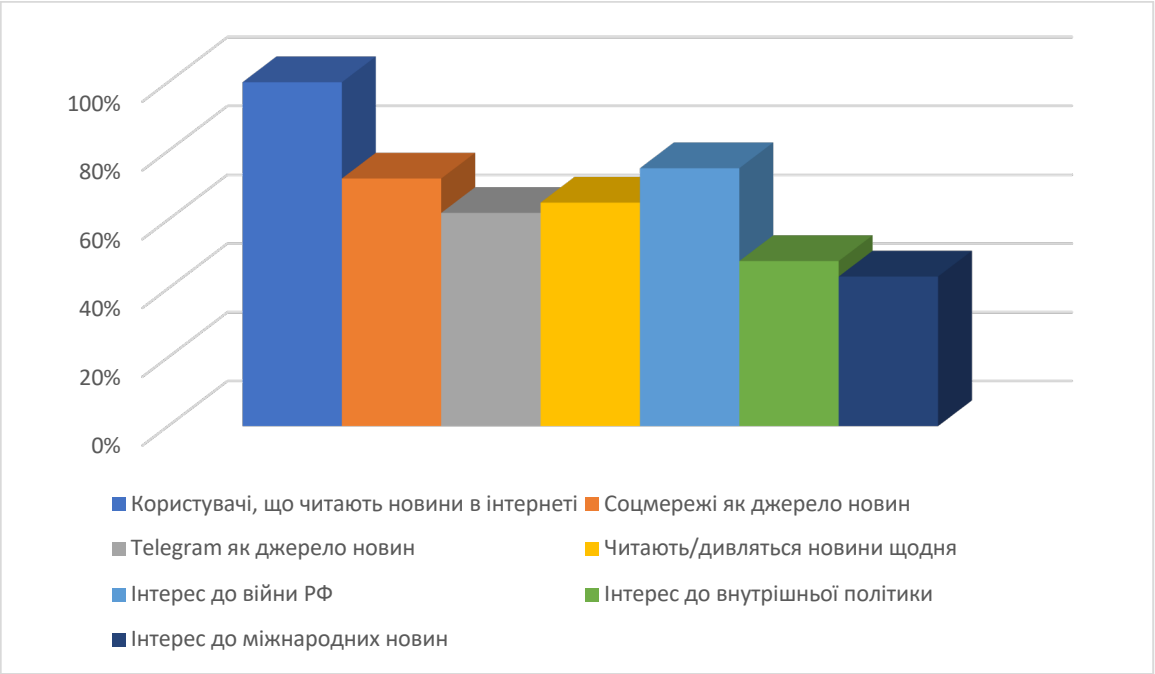

Рис. 1.4. Статистика споживання новин українцями та інтерес до тем за даними [2]

1.1. Психологічні аспекти інформаційного впливу

Психологічні аспекти інформаційного впливу становлять ключовий компонент сучасних гібридних, когнітивних і кібероперацій, оскільки саме через психіку людини відбувається перетворення зовнішніх інформаційних сигналів у внутрішні смисли, рішення та конкретні поведінкові дії. В умовах цифровізації та надмірного інформаційного навантаження психіка стає мішенню для цілеспрямованих маніпуляцій, які застосовуються як державними, так і недержавними акторами. Будь-яке повідомлення — текстове, аудіовізуальне чи інтерактивне — впливає не лише на раціональні процеси мислення, але й на емоційні реакції, мотиваційну сферу, а також на підсвідомі когнітивні автоматизми. Таким чином формується комплексна система впливу, що визначає ставлення, оцінки, вибір, готовність до дії та соціальну поведінку.

У сучасних умовах інформація дедалі частіше подається у вигляді гібридних мультимедійних потоків, які базуються на поєднанні раціональних і емоційних

тригерів, значно підвищуючи її психологічну ефективність. Цей процес підтримують алгоритми персоналізованої подачі контенту, здатні посилювати вплив через «фільтр бульбашок» та ефект когнітивного резонансу, коли користувач отримує саме ті повідомлення, які узгоджуються з його очікуваннями або підсилюють попередні переконання. У результаті інформація не лише передає зміст, але й безпосередньо конструює індивідуальну картину світу, визначаючи спосіб інтерпретації подій і тенденцій.

Окрему важливість становить вплив на емоційні процеси. Дослідження показують, що емоційно забарвлені повідомлення, зокрема ті, що апелюють до страху, гніву чи невизначеності, значно швидше поширюються в соціальних мережах і мають вищий рівень запам’ятовування. Це зумовлено специфікою роботи лімбічної системи, яка реагує швидше за когнітивну кору, сприяючи формуванню імпульсивних рішень. В інформаційних операціях емоційні тригери використовуються з метою різкого підвищення рівня тривожності, створення ефекту загрози або мобілізації певних соціальних груп на основі відчуття небезпеки та невідкладності.

Мотиваційні процеси також є вразливою сферою, оскільки інформаційний вплив здатний змінювати ціннісні орієнтири, соціальні очікування та поведінкові наміри. Наприклад, маніпулятивні меседжі можуть стимулювати відчуття провини, відповідальності, зобов'язання або, навпаки, відчуження та апатію. У кіберпросторі цей вплив посилюється через ефект анонімності, відсутність негайних соціальних санкцій та специфіку цифрового середовища, яке створює ілюзію дистанціювання від наслідків власних дій.

У сукупності всі три канали — когнітивний, емоційний та мотиваційний — формують структуру психологічної вразливості, яку активно експлуатують актори інформаційних операцій. Тому системний аналіз цих механізмів є ключовим завданням у дослідженні інформаційної безпеки, оскільки дозволяє виявити як індивідуальні, так і групові «точки входу» деструктивного впливу. Узагальнення основних механізмів наочно представлено у таблиці 1.1, що демонструє типові когнітивні, емоційні та поведінкові моделі, які стають об’єктами маніпулятивного впливу.

Таблиця 1.1.

Основні психологічні механізми інформаційного впливу

Механізм	***Суть***	***Типові прояви у цифровому середовищі***
Когнітивні викривлення	Систематичні помилки мислення, що впливають на сприйняття інформації	Ефект підтвердження; орієнтація на попередні переконання; селективне сприйняття новин
Соціальний доказ	Людина сприймає інформацію як достовірну, якщо її підтримує група	Вірусні дописи в соцмережах, «ботоферми», масові репости
Емоційний тиск	Створення афективних реакцій (страх, провина, терміновість)	Фейкові новини про небезпеку, SMS-шахрайство, повідомлення про «термінові проблеми»
Ефект авторитету	Люди схильні довіряти джерелу, яке сприймається як компетентне	Підроблені акаунти державних установ, підміна офіційних повідомлень
Ефект дефіциту	Обмеженість ресурсу або часу підвищує імпульсивність дій	«Обмежена кількість місць», «останні хвилини для реєстрації», фішинг-кампанії

Когнітивні механізми впливу та роль маніпулятивних повідомлень

Когнітивний вимір інформаційного впливу охоплює процеси сприйняття, уваги, інтерпретації та оцінювання даних. Результати психолінгвістичних досліджень свідчать, що маніпулятивні меседжі цілеспрямовано конструюються так, щоб задіяти типові когнітивні спотворення. Такі спотворення — як-от ефект підтвердження (confirmation bias), евристика доступності, ефект фреймінгу, — знижують критичність мислення та сприяють прийняттю неправдивих тверджень як достовірних.

Мова соціальної інженерії, зокрема фішингових повідомлень, дезінформаційних постів чи маніпулятивних статей, містить лексичні та синтаксичні маркери, здатні підсилювати емоційну значущість і створювати враження невідкладності. Психолінгвістичні аналізи доводять, що такі повідомлення орієнтовані на:

– спрощення складних явищ, що створює ілюзію повного розуміння;

– створення причинно-наслідкових зв'язків, яких насправді не існує;

– підкріплення вже наявних переконань, що підсилює впевненість у правильності отриманих даних;

– запровадження "ворожого образу", що стимулює поляризацію.

У країнах, які перебувають під активним інформаційним тиском, зокрема в Україні, зафіксовано систематичне використання маніпулятивних методів впливу на психіку населення. Інформаційні атаки спрямовуються на зниження довіри до державних інститутів, формування відчуття тривоги, невизначеності та зміни ціннісних орієнтирів, що створює умови для дестабілізації соціально-політичного порядку.

Емоційні канали впливу: страх, тривога, терміновість

Емоційні процеси виступають каталізатором інформаційного впливу. Емоційно заряджені повідомлення поширюються в цифровому середовищі значно швидше, ніж нейтральні, оскільки активують базові механізми виживання та соціальної взаємодії.

Однією з найбільш ефективних тактик маніпуляції є створення емоційного тиску через:

– страх (загроза, небезпека, катастрофічні сценарії);

– тривогу та невизначеність (брак чіткої інформації, суперечливі повідомлення);

– ефект терміновості (негайна необхідність реагування);

– апеляцію до авторитету (посилання на вигаданих експертів або «офіційні» джерела).

Прикладом цього є феномен «інформаційної паніки», який, за даними [4], виникає внаслідок масового поширення фейкових новин. Під впливом панічних повідомлень громадяни втрачають здатність до раціональної оцінки ситуації, що істотно підвищує ефективність кібер- та психологічних операцій противника. Панічні реакції спостерігались під час епідеміологічних криз, загрозливих політичних подій та інформаційних кампаній, спрямованих на дестабілізацію суспільства.

Мотиваційні механізми та стимулювання поведінкових реакцій

Інформаційний вплив спрямований не лише на зміну уявлень, а й на мотивацію конкретних дій: переходу за посиланням, поширення новини, здійснення фінансової операції або зміни соціальної поведінки. Мотиваційні

механізми експлуатують дефіцит уваги, прагнення до швидких рішень, бажання соціальної підтримки або належності до групи.

Особливо ефективними є тактики, які використовують:

– дефіцит (scarcity) — «акція діє лише 10 хвилин»;

– соціальний доказ — «усі вже зробили цей крок»;

– authority bias — «відомий експерт рекомендує…»;

– urgency bias — «негайно оновіть пароль».

Такі техніки широко застосовуються у фішингових атаках, маніпуляціях у соцмережах та інформаційних операціях держав-агресорів.

Вікові особливості сприйняття інформаційного впливу

Суттєві відмінності у психологічній стійкості до інформаційного впливу спостерігаються між віковими групами. За даними [5], молоді люди віком 18–24 років є найбільш вразливою категорією, оскільки вони:

– більше часу проводять у цифровому середовищі;

– схильні довіряти візуальному контенту;

– частіше реагують імпульсивно;

– мають слабший досвід розпізнавання інформаційних маніпуляцій;

– активно використовують соцмережі як основне джерело новин.

Статистично молодь піддається атакам соціальної інженерії приблизно на 20% частіше, ніж старші групи. Це визначає потребу в спеціалізованих програмах підвищення кіберобізнаності та медіаграмотності саме для цієї вікової категорії.

Психологічний вплив як інструмент гібридних операцій

У сучасних гібридних конфліктах інформаційно-психологічний вплив набуває стратегічного значення. Напади в інформаційному просторі нерідко передують технічним кібероперація̈м або супроводжують їх, оскільки підготовлене психологічне середовище полегшує реалізацію тактичних цілей. Знижена довіра, деморалізація, внутрішня поляризація та панічні настрої роблять суспільство вразливим до зовнішніх атак.

Інформаційні операції здатні:

– підривати довіру до державних інститутів;

– створювати конфлікти між соціальними групами;

– маніпулювати громадською думкою;

– послаблювати соціальну згуртованість;

– стимулювати деструктивні форми поведінки.

Таким чином, психологічний компонент становить невід'ємну частину комплексних загроз інформаційній безпеці.

Комплексна протидія психологічному впливу

З огляду на складний характер психологічних маніпуляцій ефективна протидія вимагає поєднання технічних, організаційних і поведінкових методів.

До технічних засобів належать:

– антивірусні рішення;

– багатофакторна автентифікація;

– системи фільтрації та аналізу контенту;

– виявлення аномальної поведінки користувачів.

Поведінкові методи включають:

– тренінги щодо розпізнавання соціальної інженерії;

– симуляції фішингових атак;

– тестування реакцій співробітників;

– розроблення корпоративних стандартів інформаційної безпеки.

Освітні програми з медіаграмотності мають особливе значення, оскільки вони пояснюють механізми психологічного впливу, зокрема експлуатацію страху, авторитету, терміновості та дефіциту. Візуалізація ризиків, наприклад за допомогою графіків та гістограм частоти атак, підвищує рівень усвідомлення та формує культуру безпеки як на рівні організацій, так і на рівні широкого суспільства.

1.2. Соціальні наслідки впливу Інтернету

Активне проникнення інформаційно-комунікаційних технологій у повсякденне життя зумовило суттєву трансформацію соціальної взаємодії, структури комунікацій та механізмів формування суспільної довіри. Інтернет, який упродовж останніх десятиліть став універсальним середовищем для політичних, економічних, культурних та міжособистісних процесів, формує як

нові умови соціальної інтеграції, так і нові ризики соціальної фрагментації. Сучасні дослідження доводять, що масштабність цифровізації істотно впливає на соціальну поведінку, і цей вплив має багатовимірний характер, охоплюючи як об’єктивні характеристики соціальних структур, так і суб’єктивні уявлення індивідів про власне місце у суспільстві.

Вплив Інтернету на суспільну довіру та відчуття соціальної справедливості

Одним із ключових соціальних наслідків поширення Інтернету є зниження рівня соціальної довіри, що підтверджується результатами Chinese General Social Survey [6]. Дані емпіричного дослідження засвідчують, що інтенсивне використання Інтернету має статистично значущий негативний прогнозний вплив на суспільну довіру, яка традиційно розглядається як фундаментальний чинник стабільності демократичних режимів, ефективності державного управління та соціальної згуртованості.

Медіаційним механізмом цього впливу виступає відчуття соціальної справедливості. Згідно з результатами аналізу, активна Інтернет-активність користувачів знижує суб’єктивне уявлення про справедливість суспільних відносин, сприяючи порівнянню себе з іншими, посиленню відчуття нерівності, несправедливого розподілу ресурсів і можливостей. Саме це зниження суб’єктивного почуття справедливості опосередковує зменшення рівня міжособистісної довіри. Таким чином, цифрове інформаційне середовище формує не лише поведінкові патерни індивідів, а й трансформує соціальні уявлення про справедливість, статус і взаємодію, впливаючи на соціальну тканину в довгостроковій перспективі.

Дослідники зазначають, що у мережевому просторі індивіди значно частіше порівнюють себе з іншими, що може спричиняти ефект «соціального знецінення», коли власні досягнення або соціальний статус здаються менш значущими на тлі демонстративного онлайн-успіху інших користувачів. Цей механізм створює дисонанс між об’єктивним і суб’єктивним соціальним статусом, що відповідно до дослідження [6] підвищує рівень соціального незадоволення та може формувати підґрунтя для емоційної напруги, розчарування або навіть протестної поведінки в суспільстві.

Соціальна нерівність та цифрове середовище

Окремого значення набуває вплив Інтернету на відтворення та посилення соціальних нерівностей. Доступ до інформації, цифрової інфраструктури, освітніх ресурсів і сучасних комунікаційних платформ є нерівномірним у різних соціальних групах та регіонах. Це формує феномен «цифрового розриву», коли користувачі з різними рівнями економічних і культурних ресурсів отримують нерівні можливості щодо участі у соціальних процесах.

Дослідження вказують, що цифровий розрив проявляється не лише в доступі, а й у якості та характері використання Інтернету. Особи з вищим рівнем освіти схильні використовувати мережу для роботи, навчання чи професійного розвитку, тоді як соціально вразливі групи — переважно для розваг або пасивного споживання контенту, що не сприяє соціальному піднесенню. Це, у свою чергу, посилює інституційну недовіру, формує відчуття соціального виключення та знижує можливості соціальної мобільності.

Цифрова соціальна ізоляція та зміна форм міжособистісної взаємодії

Суттєвим соціальним аспектом впливу Інтернету є зміна характеру міжособистісних зв'язків. Огляд наукової літератури демонструє, що цифрові комунікації можуть як сприяти соціальній інтеграції, так і навпаки — посилювати ізоляцію залежно від цілей та моделей користування Інтернетом.

Зокрема, соціальна взаємодія у цифрових середовищах здатна компенсувати нестачу офлайн-контактів, надаючи можливість для підтримання ширших соціальних мереж. Проте, як зазначено у джерелі [7], ефект є амбівалентним:

– Якщо Інтернет використовується для підтримки існуючих соціальних зв'язків (родинних, дружніх, професійних), то його вплив переважно позитивний: він розширює комунікативні можливості, дозволяє зберігати соціальний капітал та підвищує суб'єктивне відчуття підтримки.

– Якщо ж цифрове середовище використовується як засіб уникнення прямої соціальної взаємодії або заміни офлайн-контактів, то зростає ризик соціальної ізоляції, самотності та психологічної вразливості.

Особливо чутливою групою є молодь, яка часто перебуває у стадії формування соціальних навичок та ідентичності. Надмірне занурення у

віртуальну комунікацію може призводити до зниження якості міжособистісної взаємодії, погіршення комунікативних умінь і формування поверхневих контактів, що не забезпечують глибокої соціальної підтримки. Додатково фіксуються негативні кореляції між тривалістю онлайн-перебування та рівнем задоволеності реальними соціальними зв'язками.

Соціальна напруженість та конфліктність у цифровому середовищі

Інтернет значно підвищує динаміку поширення емоційно забарвленої інформації, конфліктних наративів і дезінформаційних матеріалів, що безпосередньо впливає на рівень соціальної настроєвості та конфліктності. В онлайн-середовищі частіше виникає ефект емоційного зараження, коли негативні настрої (агресія, фрустрація, роздратування) швидко поширюються через соціальні мережі.

Також спостерігається тенденція до формування «інформаційних бульбашок» — закритих середовищ, у яких користувачі взаємодіють переважно з тими, хто підтверджує їхні переконання. Це не лише зменшує здатність до критичного мислення, а й сприяє поляризації, формуванню стереотипів та посиленню міжгрупових конфліктів.

У результаті цифрове середовище стає платформою, яка здатна швидко масштабувати соціальні конфлікти, надати їм публічності, створити «ефект масовості» та підвищити суспільну напруженість.

Таким чином, соціальні наслідки впливу Інтернету є комплексними та проявляються на рівні індивіда, групи та суспільства загалом. Інтернет трансформує механізми формування соціальної довіри, впливає на оцінку справедливості та рівня соціального статусу, посилює або пом'якшує соціальну нерівність, модифікує структуру міжособистісних зв'язків і здатен підвищувати конфліктність та соціальну напруженість.

1.3. Політичні наслідки інформаційного впливу

Інформаційний вплив істотно позначається на політичній стабільності та легітимності державних інституцій: інтенсивні кампанії дезінформації

підривають довіру до органів влади, сприяють поляризації суспільства та можуть змінювати поведінку виборців. Так, маніпулятивні повідомлення і фейки не лише генерують конфлікти між соціальними групами та активізують негативні стереотипи, а й можуть прямо впливати на виборчі процеси — шляхом формування переваг певних політичних сил або деморалізації частини електорату. Внаслідок інформаційного тиску можливі блокування реформ, зниження ефективності управлінських рішень та загальна дестабілізація політичного курсу держави (див. огляд моніторингів і аналітики з питань дезінформації).

Емпіричні результати моніторингу підтверджують системність і спрямованість таких впливів. Зокрема, у звіті «Детектора медіа» (моніторинг 14–20 вересня 2020 р.) здійснено системний аналіз наративів, притаманних проросійській пропаганді в національних і регіональних медіа та Telegram-каналах восьми областей України. [8] Дослідження виявило концентрацію проросійських месиджів у низці видань і телеканалів (серед яких «Страна.ua», «Вести», MigNews, телеканали «112», NewsOne, ZIK, «Інтер») і окреслило домінуючі наративи: зображення України як «недодержави» з некомпетентним керівництвом; теза «нацистської держави» з перебільшенням радикалізації; наратив «громадянської війни», що заперечує участь РФ у конфлікті; а також ідеї зовнішнього управління та деградації державності (табл. 1.2). Аналітики підкреслювали, що навіть критика влади може трансформуватися у матеріалах у бік делегітимізації держави, а системний характер поширення таких наративів фіксувався у порівнянні з попередніми періодами.

Таблиця 1.2.

Основні проросійські наративи, їх характеристика та джерела поширення

Наратив	***Сутність / ключові месиджі***	***Типові медіа та канали поширення***	***Ефекти впливу на суспільство***
1. «Україна — недодержава»	Позиціонування України як неефективної, корумпованої та неспроможної до управління державою; наголос на	Інформаційні сайти («Страна.ua», «Вести»), телеканали («112», ZIK, NewsOne), регіональні медіа.	Делегітимізація влади, формування недовіри до державних інститутів, зниження довіри до реформ.

	економічній і політичній кризі.		
2. «Україна — нацистська держава»	Перебільшення та викривлення тем радикалізму, «насадження нацизму», підміна понять патріотизму та екстремізму.	Медіа з проросійською редакційною політикою, а також окремі Telegram-канали, орієнтовані на політичні скандали.	Поглиблення внутрішньої поляризації, формування образу України як «агресора» для зовнішньої аудиторії.
3. «Громадянська війна на Донбасі»	Заперечення участі РФ у війні, подання конфлікту як суто внутрішнього; звинувачення України у «невиконанні Мінських угод».	Інформаційні портали, телеканали, проросійські Telegram-канали; контент, що ретранслює позиції РФ.	Зміщення відповідальності за агресію, підрив міжнародної підтримки України, деморалізація суспільства.
4. «Україна під зовнішнім управлінням»	Твердження про контроль з боку США, ЄС або МВФ; нав'язування думки, що уряд нібито не ухвалює самостійних рішень.	ТБ-канали з політичною афілійованістю («Інтер», група каналів ОПЗЖ), Telegram-канали з анонімним адмініструванням.	Зниження довіри до проєвропейського курсу, дискредитація міжнародного партнерства, посилення антизахідних настроїв.
5. «Україна сама винна у втраті Криму»	Пропаганда тези про «нездатність України утримати Крим», заперечення незаконності анексії.	Інтернет-ЗМІ, Telegram-канали, блогери з проросійською позицією.	Легітимізація окупації, зниження відчуття національної єдності та історичної справедливості.
6. «Вибори сфальсифіковані / влада нелегітимна»	Поширення недовіри до виборчих інституцій, твердження про фальсифікації, зокрема на тимчасово окупованих територіях.	Телемарафони проросійських каналів (до 2021 р.), мережа політичних Telegram-каналів.	Підрив демократичних процедур, зменшення активності виборців, делегітимізація обраних органів влади.

Моніторингові дані свідчать також про механістичну повторюваність і скоординованість меседжів: типові теми (втрачена незалежність, зовнішнє управління, внутрішній конфлікт) регулярно посилюються через низку каналів — від інформаційних сайтів до телеграм-каналів — що підсилює ефект кумулятивного впливу на аудиторію. Такі системні наративи підвищують ймовірність перетворення інформаційних маніпуляцій у реальні політичні

наслідки (протести, падіння довіри, дипломатичні ускладнення), отже вимагають комплексного реагування, що поєднує моніторинг, фактчекінг і підвищення медіаграмотності.

Порівняльний аналіз показує, що більшість проросійських наративів структурно взаємопов'язані та виконують спільну функцію — послаблення української державності через делегітимізацію влади, підрив єдності суспільства та створення альтернативної «реальності», вигідної РФ. Канали поширення охоплюють як медіа з очевидною політичною афіліацією, так і мережу анонімних цифрових платформ, що дозволяє забезпечувати масштабування дезінформації та її адаптацію під різні аудиторії.

Отже ми приходимо до висновку: по-перше, системна природа дезінформаційних кампаній робить необхідним регулярний якісний моніторинг медіапростору; по-друге, виявлені наративи демонструють цілеспрямованість і політичну спрямованість впливу, що обґрунтовує важливість міжвідомчої координації і державної підтримки незалежних аналітичних і фактчекінгових ініціатив; по-третє, превентивні заходи мають поєднувати правові, технічні та освітні інструменти — від нормативного врегулювання до впровадження систем раннього виявлення й алгоритмів, що відстежують поширення проблемних наративів у реальному часі.

2. МЕТОДИ ТА ЗАСОБИ ВИЯВЛЕННЯ ІНФОРМАЦІЙНОГО ВПЛИВУ НА КОРИСТУВАЧІВ ГЛОБАЛЬНОЇ МЕРЕЖІ INTERNET

Системний аналіз наукових праць щодо оцінки контенту у глобальному інфопросторі дозволив визначити ключові методи та засоби ідентифікації інформаційних впливів для оцінки їхніх суспільних ризиків (рис. 2.1).

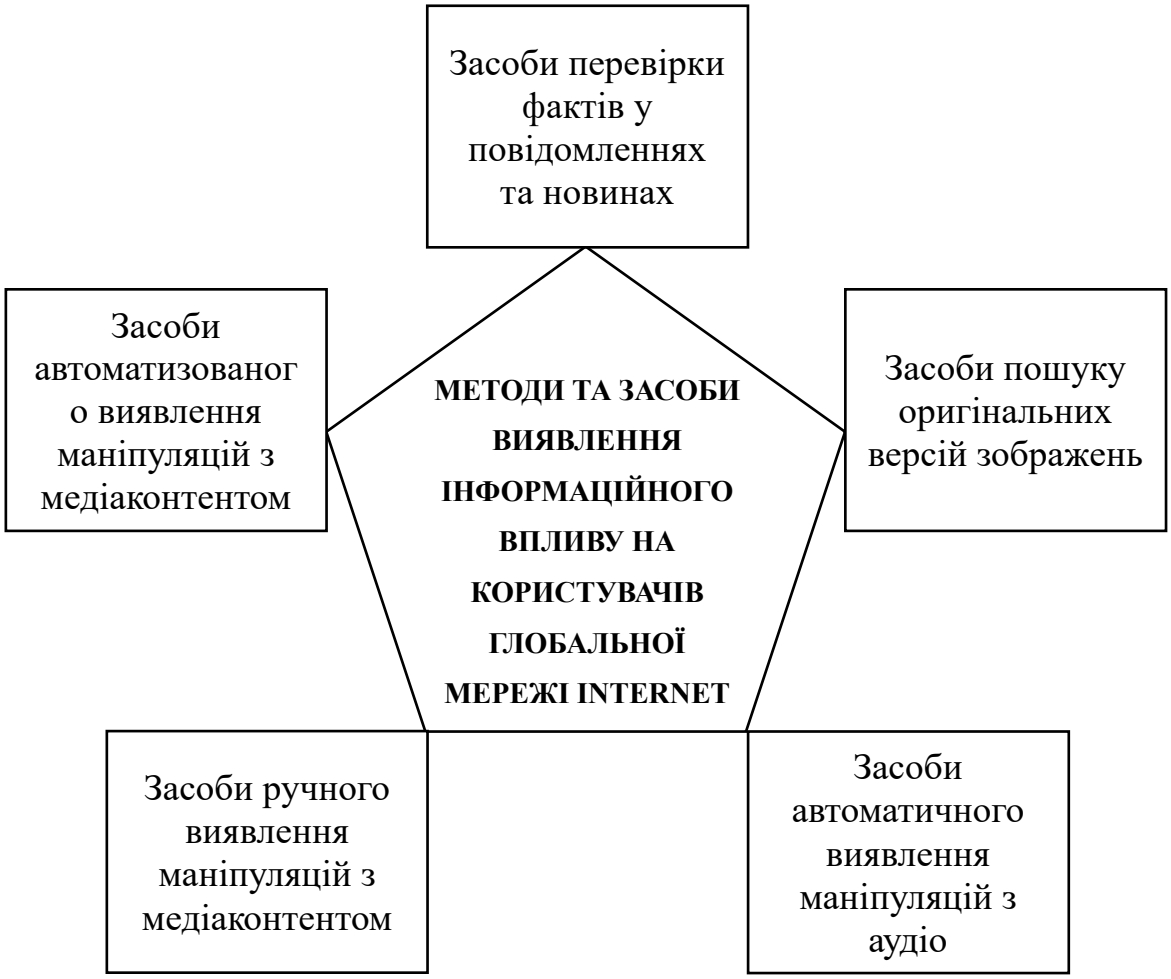

Рис. 2.1. Класифікація методів та засобів виявлення інформаційного впливу на користувачів глобальної мережі Internet

2.1. Засоби перевірки фактів у повідомленнях та новинах

FactBot (Snopes) [1] — чат-бот від сайту Snopes, що за допомогою сучасних ШІ-моделей проводить перевірку тверджень на достовірність. Він аналізує запит користувача, звертається до архівів Snopes і видає відповідь із посиланнями на підтверджені матеріали. Завдяки інтеграції з соціальними мережами можна

[1] https://www.snopes.com/factbot/

отримувати миттєві відповіді на повідомлення з фактчекінгу.

PolitiFact[2] — неупереджений фактчекінговий ресурс, який оцінює заяви політиків і публічних діячів за шкалою «Truth-O-Meter» (від «Правдиво» до «Брехливо»). PolitiFact надає архів перевірених висловлювань і допоміжні матеріали з поясненнями, що полегшують розуміння складних тем. Це дозволяє швидко знайти, чи було певне твердження перевірено раніше, за допомогою пошуку за ключовими словами та категоріями.

FactCheck.org[3] — некомерційний проєкт Університету Пенсильванії, що слідкує за точністю висловлювань у політичних дебатах, виступах і рекламних кампаніях. Сайт виступає «адвокатом» для виборців і проводить «моніторинг» фактичної точності заяв посадовців.

ChatGPT[4] — генеративна модель штучного інтелекту для ведення інтерактивного діалогу. Хоч вона й не спеціалізується на фактчекінгу, ChatGPT може допомогти з'ясувати достовірність інформації, пояснюючи складні поняття та надаючи текстові відповіді на різні запити. Завдяки API-інтеграції ChatGPT можна використовувати в чатах чи додатках для миттєвого аналізу повідомлень та створення пояснень, хоча слід перевіряти отримані відповіді через додаткові джерела.

Fake news debunker by InVID & WeVerify[5] — розширення для браузера Chrome, яке уособлює собою «універсальний інструмент» (verification «Swiss army knife») для перевірки відео та зображень. Воно значно економить час журналістів і фактчекерів при перевірці інформації у соціальних мережах. Інструмент був відзначений Інститутом Поїнтер як «один із найпотужніших інструментів для виявлення дезінформації онлайн».

Google Fact Check Explorer (toolbox.google.com)[6] — онлайн-інструмент від Google, призначений для пошуку фактчеків по всьому світу. Розробники позиціонують його як пошукову систему для фактчекінгу, яка допомагає журналістам та дослідникам дізнатися, що вже було спростовано чи

[2] https://www.politifact.com/
[3] https://www.factcheck.org/
[4] https://chatgpt.com/
[5] https://chromewebstore.google.com/detail/fake-news-debunker-by-inv/mhccpoafgdgbhnjfhkcmgknndkeenfhe
[6] https://toolbox.google.com/factcheck/explorer/search/

підтверджено. У Fact Check Explorer можна шукати перевірену інформацію за ключовими словами або навіть завантажувати зображення, щоб з'ясувати, чи були вони об'єктом фактчекінгу раніше. Результати видаються у хронологічному порядку з назвою організації та рейтингом твердження (наприклад, «некоректно», «фальшиво»).

The-Osint-Toolbox/Fact-Checking-Verification[7] — репозиторій на GitHub, який містить зібрані інструменти та ресурси для фактчекінгу і аналізу фейкових новин та зображень, згенерованих ШІ. Тут можна знайти посилання на різні онлайн-сервіси й методики перевірки інформації (наприклад, колонки перевірок, спеціалізовані браузерні розширення, інструменти для аналізу зображень тощо).

07.World[8] — персоналізована стартова сторінка (основа сайту start.me) з підбіркою важливих інформаційних ресурсів. Сторінка «07.World» містить кураторські добірки новинних джерел та інструментів фактчекінгу з усього світу, зокрема під регіоном Близького Сходу. Користувачі можуть створювати власні списки закладок з надійних ЗМІ та офіційних відомств, що дозволяє оперативно відстежувати актуальну інформацію й запобігати поширенню дезінформації.

2.2. Засоби пошуку оригінальних версій зображень

Search by Image [9] — розширення для браузерів (відкрите на GitHub) для зворотного пошуку зображень. Воно дозволяє завантажити файл або вказати URL зображення і автоматично шукати схожі чи оригінальні зображення через Google, Bing та інші пошукові системи. Наприклад, розробники цієї утиліти зазначають, що вона підтримує понад 30 пошукових платформ для порівняння фото та допомагає журналістам перевіряти автентичність візуальних матеріалів.

TinEye [10] — спеціалізована система для зворотного пошуку зображень. Після завантаження знімка або введення його адреси TinEye шукає в Інтернеті схожі копії та першоджерела фото. TinEye індексує мільярди зображень і показує, де саме вони з'являлися в мережі. Це допомагає виявити, чи не було фото

[7] https://github.com/The-Osint-Toolbox/Fact-Checking-Verification
[8] https://start.me/p/lLaoXv/07-world
[9] https://github.com/dessant/search-by-image
[10] https://tineye.com/

використано раніше в інших контекстах або чи не воно виготовлено.

Google Images [11] — сервіс Google, що дозволяє шукати картинки двома способами: за текстовим запитом і шляхом завантаження/передачі зображення (Google Lens). Можна просто ввести ключові слова для пошуку, а можна завантажити фото чи зробити знімок і знайти в мережі подібні зображення. Google Images також надає фільтри результатів за розміром, типом, кольором тощо.

Lenso.ai [12] — AI-платформа для зворотного пошуку зображень. Система використовує нейромережі для аналізу зображення за категоріями (обличчя, місця, об'єкти, стиль тощо) та генерує відповідні результати. Lenso.ai особливо сильна у пошуку точних дублікатів і визначенні першоджерел фото. Інструмент дозволяє фільтрувати результати за доменом чи ключовими словами, що полегшує перевірку походження зображення.

VisualOrigins Detector[13] — онлайн-інструмент, створений Henk van Ess (Digital Digging), який допомагає знайти найпершу публікацію зображення в Інтернеті. Він дозволяє ввести URL або завантажити файл зображення і визначає, коли це зображення з'явилося вперше. Таким чином можна відстежити першоджерела фото і з'ясувати, чи не використовувалося зображення раніше в інших контекстах.

2.3. Засоби автоматизованого виявлення маніпуляцій з медіаконтентом

Decopy.ai [14] — це вебзастосунок для виявлення зображень, створених або змінених за допомогою штучного інтелекту. Користувач завантажує фото у будь-якому популярному форматі (JPEG, PNG тощо), після чого система аналізує його структуру, пікселі та метадані, щоб визначити, чи є характерні ознаки генерації нейромережею. Інструмент орієнтований на розпізнавання контенту, створеного за допомогою таких моделей, як Midjourney, DALL-E або Stable Diffusion.

---

[11] https://www.google.com/imghp
[12] https://lenso.ai/
[13] https://visualorigins.digitaldigging.org/
[14] https://decopy.ai/ai-image-detector/

Undetectable.ai [15] — вебзастосунок, призначений для визначення, чи було зображення створене або відредаговане за допомогою штучного інтелекту. Користувач може завантажити будь-яке фото, після чого система аналізує його структуру, текстури, кольорові переходи та метадані, щоб виявити характерні сліди генерації нейромережею.

Illuminarty [16] — онлайн-інструмент, що використовує ШІ для аналізу фото. Він спроможний оцінити ймовірність штучного походження зображення і навіть вказати, за яким алгоритмом воно було згенероване. Наприклад, сервіс «дає змогу дізнатися ймовірність AI-генерації» файлу і вказує найбільш імовірні методи створення. Illuminarty доступний через вебінтерфейс і API, тому ним можна інтегрувати в системи автоматичного модераторства контенту.

AI Detect Content [17] — безкоштовний веб-інструмент, що аналізує завантажені фото та видає процентну оцінку їх штучного походження. Він спеціалізується на виявленні зображень, створених або змінених генеративними моделями (наприклад, Stable Diffusion, DALL·E, Midjourney). Після завантаження фото сервіс використовує алгоритми верифікації, аби визначити, чи було воно створено ШІ, і показує результат.

WasItAI [18] — застосунок, призначений для швидкого аналізу фото на наявність ознак штучного створення. Завантаживши зображення, користувач отримує висновок про те, чи воно є реальним чи згенерованим ШІ. За словами розробників, система порівнює візуальні особливості фото з базою реальних та синтетичних зображень, щоб «точно визначити походження». WasItAI працює у режимі онлайн і призначений допомогти користувачам відрізняти фейки від оригіналів.

2.4. Засоби ручного виявлення маніпуляцій з медіаконтентом

Forensically (29a.ch) [19] — безкоштовний онлайн-набір інструментів для

[15] https://undetectable.ai/ai-image-detector
[16] https://illuminarty.ai/en/
[17] https://aidetectcontent.com/ai-generated-image-detector/
[18] https://wasitai.com/
[19] https://29a.ch/photo-forensics/#forensic-magnifier

судової експертизи зображень. Він включає «збільшувач пікселів» (Forensic Magnifier), який дозволяє детально роздивлятися окремі фрагменти фото; Error Level Analysis (ELA) для виявлення артефактів стиснення, що вказують на правки; а також пошук повторюваних (скопійованих) фрагментів і аналіз шумів зображення. Завдяки комбінації цих засобів Forensically допомагає вручну знайти області зображення, що могли бути відредаговані чи підроблені.

FotoForensics [20] — популярна платформа для аналізу автентичності зображень. Найвідомішою її функцією є ELA-аналіз — візуалізація областей з різним рівнем стиснення JPEG, що часто вказує на редагування. Крім того, FotoForensics дозволяє переглядати EXIF-метадані (час, місце і пристрій зйомки) і порівнювати зображення з його версією з бази даних, щоб знайти відмінності. Такий інструмент часто використовують фактчекери для ручної перевірки підозрілих фото.

Fake Image Detector [21] — вебсервіс для швидкої перевірки достовірності зображень. Він використовує як метадані, так і ELA-аналітику для фіксації маніпуляцій. Наприклад, на головній сторінці Fake Image Detector зазначено, що сервіс «ідентифікує змінені зображення за допомогою аналізу метаданих та ELA». Користувачі можуть завантажити знімок, після чого система проводить кілька перевірок і показує відмічені ризикові області. Такий інструмент простий у використанні для стартової оцінки правдоподібності фото.

2.5. Засоби автоматичного виявлення маніпуляцій з аудіо

Fake Voice Detection [22] — відкритий проєкт від Dessa (тепер частина OpenAI), призначений для виявлення підроблених голосових записів, створених за допомогою нейромереж. Алгоритм аналізує спектрограми аудіо та шукає ознаки синтетичного звучання, характерні для генеративних моделей. Розробка була спрямована на протидію «deepfake»-аудіо, що може використовуватись для маніпуляцій чи шахрайства.

---

[20] https://fotoforensics.com/
[21] https://www.fakeimagedetector.com/
[22] https://github.com/dessa-oss/fake-voice-detection

Resemblyzer [23] — бібліотека Python, розроблена командою Resemble AI, яка обчислює векторні представлення (ембеддинги) голосів. Вона використовується для визначення подібності між аудіозаписами — наприклад, для перевірки, чи належать два записи одній людині, або для виявлення підробок. Resemblyzer широко застосовується у проєктах, пов'язаних із розпізнаванням мовця, класифікацією голосів і детекцією синтетичної мови.

AI Voice Detector [24] — онлайн-сервіс для швидкої перевірки, чи є аудіо створеним штучним інтелектом. Користувач завантажує запис або вставляє посилання, після чого система оцінює його на предмет deepfake-ознаків. Алгоритм аналізує частотний спектр, паузи, інтонації та цифрові артефакти, що відрізняють людську мову від синтетичної. Результат подається у відсотках — як «ймовірність, що голос створений ШІ».

Hiya Deepfake Voice Detector [25] — розширення для браузера Chrome, яке автоматично аналізує онлайн-аудіо (наприклад, у відео чи дзвінках) і визначає, чи містить воно ознаки синтетичної генерації. Hiya використовує моделі машинного навчання для виявлення нехарактерних спектральних патернів і шумів, притаманних ШІ-голосам. Зручність полягає в тому, що перевірка відбувається безпосередньо під час відтворення контенту.

ElevenLabs AI Speech Classifier [26] — інструмент від компанії ElevenLabs, який дозволяє визначити, чи був голосовий запис створений за допомогою ШІ. Сервіс аналізує аудіо, порівнюючи його характеристики зі зразками, створеними власною системою генерації голосів ElevenLabs. Користувач отримує відсоткову оцінку ймовірності, що запис є синтетичним.

[23] https://github.com/resemble-ai/Resemblyzer
[24] https://aivoicedetector.com/
[25] https://chromewebstore.google.com/detail/hiya-deepfake-voice-detec/akmieeldmgcllmokbpaibfelofjiilpc
[26] https://elevenlabs.io/ai-speech-classifier

3. СЦЕНАРІЇ ПРАКТИЧНОЇ ОЦІНКИ ЗАСОБІВ ВИЯВЛЕННЯ ІНФОРМАЦІЙНОГО ВПЛИВУ

3.1. Сценарій оцінки засобів перевірки фактів у повідомленнях та новинах

Мета: оцінити здатність інструментів перевірки фактів визначати достовірність новинних статей за допомогою ключових слів, згенерованих штучним інтелектом.

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

1. Йохауг, Ріібер, Айзенбіхлер та інші. Зірки лижного спорту, які завершили кар'єру в 2025 році [9] — **TRUE**
2. General Motors допомагає ветеранам знайти нове життя після служби [10] — **TRUE**
3. Солдаты ВСУ начали сдаваться от голода и жажды: запертым в котлах под Покровском и Купянском отвели две недели [11] — **FALSE**
4. Глава города Дрогобыч Львовской области лишил город Буффало, штат Нью-Йорк, звания города-побратима [12] — **FALSE**

Кроки

1. Вибрати одну новину зі списку.
2. На основі тексту цієї новини сформувати набір релевантних ключових слів. Для цього використати модель ChatGPT із запитом на кшталт: «Привіт. Будь ласка, створи список найбільш релевантних ключових слів для цієї статті».
3. Ввести отримані ключові слова в обраний інструмент перевірки фактів.
4. Зафіксувати результат оцінювання, який видасть інструмент.

Очікувані результати: інструменти перевірки фактів повинні відповідати початковій класифікації новини: якщо стаття позначена «FALSE», то інструмент має розпізнати її як неправдиву або ненадійну, а якщо «TRUE» — як достовірну.

Результати: результати оцінювання кожного інструменту звести у таблицю.

Подальша обробка: відсортувати інструменти за отриманими оцінками.

3.2. Сценарій оцінки засобів пошуку оригінальних версій зображень

Мета: оцінити ефективність інструментів пошуку зображень для знаходження конкретних фото в Інтернеті.

Вхід: два набори зображень:

- 15 оригінальних нерепереоблених фотографій (існуючі) — https://tinyurl.com/ybsjz6nw
- 15 зображень, штучно згенерованих (створені ШІ) — https://tinyurl.com/nkud6a2y

Кроки

1. Вибрати зображення з одного з наборів.
2. Використати інструмент зворотного пошуку зображень, щоб знайти в Інтернеті відповідну оригінальну версію цього фото.
3. Зафіксувати результат: чи вдалося інструменту знайти відповідне зображення.

Очікувані результати: інструменти мають знаходити реально існуючі зображення (позначені як «існують») і не знаходити ті, що позначені як «не існують».

Результати: зібрати результати оцінювання кожного інструменту в таблицю.

Подальша обробка: відсортувати інструменти за отриманими оцінками.

3.3. Сценарій оцінки автоматизованих засобів виявлення маніпуляцій з медіа-контентом

Мета: оцінити ефективність інструментів автоматичної перевірки зображень на предмет того, чи були вони створені або змінені за допомогою штучного інтелекту.

Вхід: ті ж набори зображень, що й у попередньому сценарії:

- 15 оригінальних нерепереоблених фотографій (існуючі) — https://tinyurl.com/ybsjz6nw
- 15 зображень, штучно згенерованих (створені ШІ) — https://tinyurl.com/nkud6a2y

Кроки

1. Вибрати зображення з набору.
2. Використати обраний інструмент для автоматичного виявлення маніпуляцій (наприклад, аналіз або перевірку на ознаки генерації ШІ).
3. Зафіксувати результат класифікації – чи зображення є «оригінальним» або «згенерованим ШІ».

Очікувані результати: інструменти повинні правильно ідентифікувати тип зображення: «існуюче» (оригінал) чи «створене ШІ».

Результати: зібрати результати роботи кожного інструменту в таблицю.

Подальша обробка: відсортувати інструменти за їхніми оцінками.

3.4. Сценарій оцінки засобів ручного виявлення маніпуляцій з медіа-контентом

Мета: оцінити ефективність інструментів, які дозволяють вручну аналізувати фотографії для виявлення змін або обробки.

Вхід: два набори зображень:

- 15 оригінальних, непереоблених фотографій (оригінальні) https://tinyurl.com/ybsjz6nw
- 15 змінених, ретушованих або фотошоплених фотографій (змінені) — https://tinyurl.com/mr47czja

Кроки

1. Вибрати зображення з одного з наборів.
2. Використати інструмент для ручного аналізу зображення, застосовуючи його вбудовані функції (наприклад, перегляд метаданих EXIF, аналіз колірних гістограм тощо).
3. Зафіксувати результат класифікації (зображення є оригіналом чи зміненим) та описати виявлені типи маніпуляцій.

Очікувані результати: інструменти мають надати аналітику можливість визначити, які саме зміни були внесені у зображення, щоб правильно класифікувати його як «оригінал» або «змінене».

Результати: узагальнити результати оцінювання кожного інструменту в

таблиці.

Подальша обробка: відсортувати інструменти за отриманими оцінками.

3.5. Сценарій оцінки засобів автоматичного виявлення маніпуляцій з аудіо

Мета: оцінити ефективність інструментів, які автоматично визначають штучно згенеровані або змінені голосові записи (deepfake-аудіо), використовуючи алгоритми машинного навчання, спектральний аналіз і перевірку метаданих.

Вхід: два набори аудіозаписів:

- 15 оригінальних людських голосових записів — https://tinyurl.com/mpp5h6xu
- 15 згенерованих за допомогою ШІ голосових записів — https://tinyurl.com/2r24tvhj

Кроки

1. Обрати один аудіозапис із будь-якої з двох папок.
2. Завантажити файл у відповідний інструмент для автоматичного аналізу
3. Виконати аналіз за допомогою доступних функцій сервісу (класифікація оригінальності, відображення спектрограм, визначення рівня ймовірності синтетичного походження тощо).
4. Зафіксувати результат автоматичної класифікації (аудіо — «оригінал» чи «змінене») і, якщо доступно, зазначити відсоткову ймовірність підробки.

Очікувані результати: інструменти мають автоматично визначити, чи є аудіозапис автентичним або створеним/зміненим за допомогою ШІ, та надати оцінку достовірності у відсотках або категоріях.

Результати: узагальнити отримані оцінки ефективності кожного інструменту в таблиці (параметри — точність класифікації, швидкість аналізу, наявність детальної звітності, зручність використання).

Подальша обробка: відсортувати інструменти за отриманими середніми оцінками ефективності, виділити найточніші системи для подальшого використання у дослідженні.

4. РЕЗУЛЬТАТИ ПРАКТИЧНОЇ ОЦІНКИ ЗАСОБІВ ВИЯВЛЕННЯ ІНФОРМАЦІЙНОГО ВПЛИВУ

У межах дослідження здійснено практичну оцінку інструментів для виявлення інформаційного впливу, що дало змогу систематизувати наявні засоби, оцінити їхню ефективність і сформулювати рекомендації щодо подальшого застосування. Отримані результати засвідчили, що більшість сучасних рішень, призначених для виявлення маніпуляцій з даними, характеризуються достатньою точністю та функціональністю під час розв'язання завдань.

4.1. Засоби перевірки фактів у повідомленнях та новинах

Таблиця оцінювання інструментів за визначеною методикою доступна за посиланням https://tinyurl.com/5dnpw32t.

Так, ChatGPT показав високу точність із 75% правильних оцінок (TRUE), тоді як FactCheck.org та PolitiFact мали значну частку випадків, коли оцінку надати було неможливо (UNDEFINED — понад 60%). Інші сервіси, такі як Fake news debunker by InVID & WeVerify та The-Osint-Toolbox/Fact-Checking-Verification, також досягли 75% правильних оцінок (табл 4.1.), але частина оцінок залишалася невизначеною. Менш ефективними виявилися 07.World, FactBot і Google Fact Check Explorer, які часто не могли надати оцінку або давали низьку точність.

Таблиця 4.1. Результати оцінки інструментів перевірки фактів

Засіб	Відсоток відповідної (TRUE), невідповідної (FALSE) оцінки фактів та неможливість (UNDEFINED) надати оцінку		
	TRUE	FALSE	UNDEFINED
ChatGPT	75,0%	25,0%	0,0%
Fake news debunker by InVID & WeVerify	75,0%	0,0%	25,0%

Засіб	Відсоток відповідної (TRUE), невідповідної (FALSE) оцінки фактів та неможливість (UNDEFINED) надати оцінку		
	TRUE	FALSE	UNDEFINED
The-Osint-Toolbox/Fact-Checking-Verification	75,0%	0,0%	25,0%
Google Fact Check Explorer	50,0%	0,0%	50,0%
07.World	27,5%	0,0%	72,5%
FactCheck.org	20,0%	17,5%	62,5%
FactBot (Snopes)	17,5%	17,5%	65,0%
PolitiFact	15,0%	22,5%	62,5%

4.2. Засоби пошуку оригінальних версій зображень

Таблиця оцінювання інструментів за визначеною методикою доступна за посиланням https://tinyurl.com/y6hca52v.

Інструменти для зворотного пошуку зображень продемонстрували обмежену точність у верифікації медіафайлів (табл. 4.2.). TinEye показав 33,3% правильних оцінок, тоді як інші сервіси, такі як Lenso.ai, Google Images, Visual Origins Detector та Search by Image, мали ще нижчий показник точності (від 16,7% до 26,7%).

Таблиця 4.2. Результати оцінки інструментів пошуку оригінальних версій зображень

Засіб	Відсоток відповідної (TRUE), невідповідної (FALSE) оцінки фактів та неможливість (UNDEFINED) надати оцінку		
	TRUE	FALSE	UNDEFINED
TinEye	33,3%	66,7%	0,0%

Засіб	Відсоток відповідної (TRUE), невідповідної (FALSE) оцінки фактів та неможливість (UNDEFINED) надати оцінку		
	TRUE	FALSE	UNDEFINED
Lenso.ai	26,7%	73,3%	0,0%
Google Images	23,3%	76,7%	0,0%
Visual Origins Detector	20,0%	80,0%	0,0%
Search by Image	16,7%	83,3%	0,0%

4.3. Засоби автоматизованого виявлення маніпуляцій з медіаконтентом

Таблиця оцінювання інструментів за визначеною методикою доступна за посиланням https://tinyurl.com/b6mv46tm.

Інструменти для виявлення контенту, створеного штучним інтелектом, показали високу точність оцінки фактів (рис 4.3.). Найефективнішим виявився WasItAI із 93,3% правильних оцінок (TRUE), тоді як Undetectable.ai та Decopy.ai досягли 76,7% і 73,3% відповідно. Illuminarty показав 66,7% точності, а AI Detect Content мав рівні показники TRUE та FALSE по 50%. Усі інструменти давали оцінку для кожного випадку (UNDEFINED = 0%), що свідчить про їхню повну придатність до використання.

Таблиця 4.3 Результати оцінки інструментів автоматизованого виявлення маніпуляцій з медіаконтентом

Засіб	Відсоток відповідної (TRUE), невідповідної (FALSE) оцінки фактів та неможливість (UNDEFINED) надати оцінку		
	TRUE	FALSE	UNDEFINED
WasItAI	93,3%	6,7%	0,0%

Засіб	Відсоток відповідної (TRUE), невідповідної (FALSE) оцінки фактів та неможливість (UNDEFINED) надати оцінку		
	TRUE	FALSE	UNDEFINED
Undetectable.ai	76,7%	23,3%	0,0%
Decopy.ai	73,3%	26,7%	0,0%
Illuminarty	66,7%	33,3%	0,0%
AI Detect Content	50,0%	50,0%	0,0%

4.4. Засоби ручного виявлення маніпуляцій з медіаконтентом

Таблиця оцінювання інструментів за визначеною методикою доступна за посиланням https://tinyurl.com/3z3k5mns.

Інструменти для аналізу та виявлення фейкових зображень продемонстрували середню точність оцінки фактів (табл. 4.4). Forensically (29a.ch) досяг 70% правильних оцінок (TRUE), FotoForensics показав 66,7%, тоді як Fake Image Detector мав лише 40% точності. В усіх випадках інструменти надавали оцінку для кожного випадку (UNDEFINED = 0%), проте частка невірних оцінок (FALSE) залишалася значною, особливо для Fake Image Detector.

Таблиця 4.4 Результати оцінки інструментів ручного виявлення маніпуляцій з медіаконтентом

Засіб	Відсоток відповідної (TRUE), невідповідної (FALSE) оцінки фактів та неможливість (UNDEFINED) надати оцінку		
	TRUE	FALSE	UNDEFINED
Forensically (29a.ch)	70,0%	30,0%	0,0%
FotoForensics	66,7%	33,3%	0,0%
Fake Image Detector	40,0%	60,0%	0,0%

4.5. Засоби автоматичного виявлення маніпуляцій з аудіо

Таблиця оцінювання інструментів за визначеною методикою доступна за посиланням https://tinyurl.com/4945wafy.

Інструменти для виявлення фейкового голосового контенту показали високу точність оцінки фактів (табл. 4.5). Найефективнішим виявився Fake Voice Detection із 93,3% правильних оцінок (TRUE), слідом йдуть AI Voice Detector (90%) та Resemblyzer (86,7%). Hiya Deepfake Voice Detector та ElevenLabs AI Speech Classifier також продемонстрували високу точність (86,7% та 83,3% відповідно).

Таблиця 4.5 Результати оцінки інструментів автоматичного виявлення маніпуляцій з аудіо

Засіб	Відсоток відповідної (TRUE), невідповідної (FALSE) оцінки фактів та неможливість (UNDEFINED) надати оцінку		
	TRUE	FALSE	UNDEFINED
Fake Voice Detection	93,3%	6,7%	0,0%
AI Voice Detector	90,0%	0,0%	10,0%
Resemblyzer	86,7%	13,3%	0,0%
Hiya Deepfake Voice Detector	86,7%	3,3%	10,0%
ElevenLabs AI Speech Classifier	83,3%	6,7%	10,0%

ВИСНОВКИ

Проведене дослідження дозволило комплексно проаналізувати природу, механізми та наслідки інформаційного впливу на користувачів глобальної мережі Інтернет, а також оцінити ефективність сучасних методів і засобів виявлення дезінформації та медіаманіпуляцій. Підсумовуючи результати роботи, можна сформулювати такі узагальнені висновки.

1. Інформаційний вплив у цифровому середовищі має комплексний когнітивно-емоційно-мотиваційний характер. Інформаційно-психологічний вплив реалізується через поєднання маніпулятивних повідомлень, цілеспрямованих тригерів та експлуатації когнітивних упереджень користувачів. Інформація у цифровому середовищі не лише передає зміст, але й формує картини світу, поведінкові патерни та соціальні орієнтири, виступаючи інструментом впливу на масову свідомість. Психологічна вразливість користувачів підсилюється надмірною інтенсивністю інформаційних потоків, алгоритмічною персоналізацією та емоційним забарвленням контенту.

2. Соціальні наслідки інтернет-впливу є системними та довготривалими. Активне споживання цифрового контенту корелює зі зниженням соціальної довіри, порушенням відчуття соціальної справедливості, формуванням інформаційних «бульбашок» та зростанням ризику соціальної ізоляції. Інтернет-середовище сприяє посиленню міжгрупових конфліктів, підвищенню соціальної фрагментації та виникненню когнітивного дисонансу, що особливо проявляється у вразливих груп населення.

3. Політичні наслідки інформаційного впливу безпосередньо пов’язані із загрозами державній безпеці та легітимності влади. Дезінформація здатна знижувати довіру до державних інститутів, дестабілізувати виборчі процеси, провокувати поляризацію та радикалізацію суспільства, а також створювати сприятливі умови для втручання зовнішніх акторів. Виявлено поширені наративи, спрямовані на делегітимізацію української державності, підрив довіри до влади та формування образу України як нестабільної держави. Це прямо впливає на національну безпеку та міжнародну суб’єктність держави.

4. Сучасні засоби виявлення інформаційного впливу забезпечують різні рівні точності, функціональності та оперативності. Засоби перевірки фактів, аналізу зображень, аудіо та мультимедійного контенту, а також ручні та автоматизовані методи виявлення маніпуляцій демонструють значну різноманітність та практичну ефективність, а саме:

– фактчекінгові платформи мають високий рівень точності, але низьку оперативність у кризових інформаційних ситуаціях;

– засоби пошуку оригінальних зображень ефективні проти реверсивних фейків, але малоефективні при зміні метаданих чи генерації AI-контенту;

– автоматизовані детектори маніпуляцій працюють добре для структурно простих підробок, проте складні deepfake-технології потребують спеціалізованих алгоритмів;

– ручні методи експерта залишаються найбільш надійними, але потребують високої кваліфікації і тривалого часу;

– інструменти виявлення аудіоманіпуляцій перебувають у стадії активного розвитку, але вже демонструють здатність виявляти цифрову реконструкцію голосу та штучні інтонаційні аномалії.

Таким чином, жоден метод не є універсальним, що вимагає комплексного підходу.

5. Розроблені сценарії практичної оцінки дозволяють створити універсальну модель тестування інструментів, забезпечують методологічну базу для визначення сильних і слабких сторін різних засобів, а також можуть бути використані для побудови систематизованої процедури виявлення дезінформації в реальних умовах. Вони можуть застосовуватися:

– у навчальному процесі;

– у практичній діяльності аналітичних підрозділів;

– у системах кіберзахисту органів державної влади;

– у медійному та журналістському середовищі.

6. Інтеграція технічних і психологічних інструментів є ключовою умовою ефективної протидії інформаційному впливу. Успішне виявлення дезінформації неможливе без синергії трьох напрямів:

Технологічного — автоматизовані системи аналізу даних, детектори фейків, алгоритми машинного навчання.

Аналітичного — експертні оцінки, контент-аналіз, когнітивні моделі оцінки ризиків.

Поведінкового та освітнього — тренінги з медіаграмотності, симуляції атак соціальної інженерії, підвищення критичного мислення.

Використання виключно технічних засобів без психологічного компонента значно знижує загальну ефективність системи кіберзахисту.

7. Для підвищення захищеності держави та суспільства необхідна багаторівнева система протидії інформаційним загрозам, яка зводиться до:

– розвитку національних інструментів автоматичного виявлення дезінформації;

– створення централізованої міжвідомчої системи аналізу інформаційних атак;

– удосконалення нормативно-правової бази у сфері інформаційної безпеки;

– підтримки фактчекінгових ініціатив та медіаінститутів;

– інтеграції міжнародних практик та стандартів НАТО і ЄС;

– розвитоку цифрової грамотності населення та спеціалізованої підготовки для фахівців.

Узагальнюючи результати дослідження, можна стверджувати, що інформаційний вплив у глобальній мережі є багатокомпонентним явищем, яке поєднує психологічні, соціальні, технологічні та політичні аспекти. Ефективна протидія вимагає не лише сучасних технічних інструментів, але й системного розвитку аналітичної, нормативної та освітньої інфраструктури. Представлений комплекс методів і сценаріїв оцінки створює основу для практичного застосування отриманих результатів у сфері кібербезпеки, державного управління, журналістики та громадянського суспільства.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. BazKhan S. GSSR — global social sciences review. GSSR — Global Social Sciences Review. URL: https://www.gssrjournal.com/article/digital-literacy-and-its-influence-on-media-consumption-habits-in-the-global-south (date of access: 10.11.2025).
2. Internet and social media users in the world 2025 | Statista. Statista. URL: https://www.statista.com/statistics/617136/digital-population-worldwide/ (date of access: 10.11.2025).
3. Мар'яна Зінченко. Українці найчастіше дізнаються новини з соцмереж, – Центр Разумкова. detector.media. URL: https://detector.media/infospace/article/239646/2025-04-02-ukraintsi-naychastishe-diznayutsya-novyny-z-sotsmerezh-tsentr-razumkova/ (дата звернення: 10.11.2025).
4. Psycholinguistic aspects of humanitarian component of cybersecurity | PSYCHOLINGUISTICS. PSYCHOLINGUISTICS. URL: https://psycholing-journal.com/index.php/journal/article/view/712 (date of access: 10.11.2025).
5. Social engineering statistics: reports 2025. WifiTalents. URL: https://wifitalents.com/social-engineering-statistics/ (date of access: 10.11.2025).
6. Internet use and social trust: empirical analysis based on CGSS2021 — PubMed. PubMed. URL: https://pubmed.ncbi.nlm.nih.gov/39830694/ (date of access: 10.11.2025).
7. Associations of internet access with social integration, wellbeing and physical activity among adults in deprived communities: evidence from a household survey — BMC Public Health. BioMed Central. URL: https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-019-7199-x (date of access: 10.11.2025).
8. ГО «Детектор медіа». Моніторинг (про)російської дезінформації в українських медіа за 14–20 вересня 2020 року. detector.media. URL: https://detector.media/propahanda_vplyvy/article/181162/2020-10-01-monitoryng-

prorosiyskoi-dezinformatsii-v-ukrainskykh-media-za-1420-veresnya-2020-roku/ (дата звернення: 11.11.2025).

9. Champion.com.ua. Йохауг, Ріібер, Айзенбіхлер та інші. Зірки лижного спорту, які завершили кар'єру в 2025 році. Champion.com.ua. URL: https://champion.com.ua/ukr/ski/12-zirok-lizhnih-vidiv-sportu-yaki-zavershili-kar-yeru-v-2025-roci-1055567 (дата звернення: 13.11.2025).

10. General Motors допомагає ветеранам знайти нове життя після служби. Auto24. URL:https://auto.24tv.ua/general_motors_dopomahaie_veteranam_znaity_nove_zhyttia_pislia_sluzhby_n67589 (дата звернення: 13.11.2025).

11. Солдаты ВСУ начали сдаваться от голода и жажды: запертым в котлах под покровском и купянском отвели две недели — KP.RU. kp.ru — Сайт «Комсомольской правды». URL: https://www.kp.ru/daily/27739/5166938/ (дата звернення: 13.11.2025).

12. Глава города Дрогобыч Львовской области лишил город Буффало, штат Нью-Йорк, звания города-побратима — Лента новостей Харькова. Лента новостей Харькова. URL: https://news-kharkov.ru/other/2025/03/07/143834.html (дата звернення: 13.11.2025).