Міністерство освіти і науки України
Київський національний університет імені Тараса Шевченка

Факультет інформаційних технологій
Кафедра кібербезпеки та захисту інформації

ДОПУСТИТИ ДО ЗАХИСТУ:
В.о. завідувача кафедри
кібербезпеки та захисту
інформації
_____ Іван ПАРХОМЕНКО
«___» червня 2025 р.

ПОЯСНЮВАЛЬНА ЗАПИСКА

кваліфікаційної роботи

галузь знань 12 Інформаційні технології
(шифр і назва галузі знань)
спеціальність 125 Кібербезпека
(код і назва спеціальності)
освітній ступень бакалавр
освітня програма Кібербезпека
(назва освітньо-професійної програми)
на тему: «Механізм автоматизованої протидії інформаційній війні»

Виконавець: студент IV курсу, групи КБ-41

_____ **Вадим ВИНОКУР**
(підпис) (ім'я, прізвище)

	Підпис	**Ім'я ПРІЗВИЩЕ**
Керівник		Сергій ДАКОВ
Нормоконтроль		Олександр ТОРОШАНКО

Київ 2025

Міністерство освіти і науки України
Київський національний університет імені Тараса Шевченка

Факультет інформаційних технологій
Кафедра кібербезпеки та захисту інформації

ЗАТВЕРДЖЕНО:
В.о. завідувача кафедри
кібербезпеки
та захисту інформації
_____ Іван ПАРХОМЕНКО
«29» листопада 2024 р.

ЗАВДАННЯ
на виконання кваліфікаційної роботи

спеціальності 125 Кібербезпека
(код і назва спеціальності)

освітньої програми Кібербезпека
(назва освітньо-професійної програми)

Студенту **КБ-41** (група) **Винокуру Вадиму Сергійовичу** (прізвище ім'я по батькові)

Тема кваліфікаційної роботи Механізм автоматизованої протидії інформаційній війні

1. ПІДСТАВИ ДЛЯ ПРОВЕДЕННЯ РОБОТИ

Тема кваліфікаційної роботи затверджена на засіданні кафедри кібербезпеки та захисту інформації протокол №6 від 28.11.2024 р.

2. ВИХІДНІ ДАНІ ДЛЯ ПРОВЕДЕННЯ РОБІТ

Концепція автоматизованого механізму протидії інформаційній війні

3. ЗМІСТ РОЗРАХУНКОВО-ПОЯСНЮВАЛЬНОЇ ЗАПИСКИ

Необхідно проаналізувати загрози інформаційної війни та існуючі методи. протидії. Розробити концептуальну модель автоматизованого механізму, Обґрунтувавши його компоненти. Здійснити програмну реалізацію прототипу та визначити перспективи.

4. ВИМОГИ ДО РЕЗУЛЬТАТІВ ВИКОНАННЯ РОБОТИ

Практична цінність Розроблена модель автоматизованого механізму, що є

Основою створення ефективніших систем протидії інформаційній війні.

5. ДАТА ВИДАЧІ ЗАВДАННЯ

Дата видачі завдання: 29 листопада 2024 року

Завдання видав _____ Сергій ДАКОВ
(підпис) (ім'я, прізвище)

Завдання прийняв до виконання _____ Вадим ВИНОКУР
(підпис) (ім'я, прізвище)

КАЛЕНДАРНИЙ ПЛАН

№ п/п	Найменування етапів робіт	Строки виконання робіт (початок-кінець)	Відмітка про виконання
1	Уточнення постановки задачі	29.11.2024 – 05.12.2024	*виконано*
2	Аналіз літератури	06.12.2024 – 29.12.2024	*виконано*
3	Обґрунтування вибору рішення	15.01.2025 – 29.01.2025	*виконано*
4	Обґрунтування методів дослідження та інструментарію розробки	30.01.2025 – 14.02.2025	*виконано*
5	Проектування архітектури та модулів автоматизованого механізму.	15.02.2025 – 28.02.2025	*виконано*
6	Практична реалізація та апробація елементів запропонованого механізму.	29.02.2025 – 29.03.2025	*виконано*
7	Розробка та тестування програмного прототипу компонента механізму.	01.04.2025 – 15.05.2025	*виконано*
8	Оформлення пояснювальної записки	16.05.2025 – 19.05.2025	*виконано*
9	Підготовка до захисту кваліфікаційної роботи	20.05.2025 – 13.06.2025	*виконано*

Завдання видав _____ Сергій ДАКОВ
(підпис) (ім'я, прізвище)

Завдання прийняв до виконання _____ Вадим ВИНОКУР
(підпис) (ім'я, прізвище)

Термін подання кваліфікаційної роботи до ЕК 13 червня 2025 року

РЕФЕРАТ

Пояснювальна записка до кваліфікаційної роботи «Механізм автоматизованої протидії інформаційній війні»: містить 75 сторінок основного тексту (від Вступу до Висновків включно), 8 рисунів, 9 таблиць. Список використаних джерел містить 31 найменування і займає 4 сторінки.

Метою роботи є підвищення ефективності протидії інформаційним загрозам шляхом дослідження існуючих методів та розробки концептуальної моделі автоматизованого механізму, спрямованого на оптимізацію процесів їх виявлення, аналізу та реагування.

Для досягнення зазначеної мети поставлено наступні завдання:

1) Проаналізувати теоретико-правові засади та сучасний стан проблеми інформаційної війни, включаючи класифікацію її загроз, роль соціальних мереж як інструменту впливу та особливості нормативного регулювання сфери інформаційної безпеки в Україні.

2) Дослідити існуючі методи та практичний досвід протидії інформаційним загрозам, проаналізувавши національні та міжнародні підходи, механізми виявлення дезінформації в онлайн-середовищі та оцінивши ефективність поточних рішень.

3) Розробити концептуальну модель автоматизованого механізму протидії інформаційним загрозам, обґрунтувавши його необхідність, визначивши ключові вимоги, архітектуру, функціональні компоненти та перспективи розвитку й інтеграції.

Об'єктом дослідження є процеси селекції і обробки інфромації під час інформаційної війни

Предметом дослідження є автоматизовані методи виявлення та протидії дезінформації, зокрема механізми аналізу, перевірки та маркування недостовірного контенту в медійному просторі.

Практична цінність роботи полягає у розробці моделі автоматизованого механізму, який може слугувати основою для створення більш ефективних інструментів протидії інформаційним загрозам

Ключові слова: інформаційна війна, протидія інформаційним загрозам, автоматизований механізм, дезінформація, пропаганда, соціальні мережі, аналіз загроз, обробка даних.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ ТА СКОРОЧЕНЬ

AI	—	Artificial intelligence (Штучний інтелект)
ШІ	—	Штучний інтелект
NLP	—	Обробка природної мови (Natural Language Processing)
МН	—	Машинне навчання
Big Data	—	Великі дані (термін для аналізу великих обсягів даних)
ЦПД	—	Центр протидії дезінформації
НАТО	—	Організація Північноатлантичного договору (North Atlantic Treaty Organization)
NATO StratCom COE	—	Центр стратегічних комунікацій НАТО (NATO Strategic Communications Centre of Excellence)
CCDCOE	—	Об’єднаний центр передових технологій з кібероборони НАТО (NATO Cooperative Cyber Defence Centre of Excellence)
WMGIC	—	Глобальний інноваційний конкурс Вільяма та Мері (William & Mary’s Global Innovation Challenge)
NAADA	—	Північноатлантичне антидезінформаційне агентство (North Atlantic Anti-Disinformation Agency)
IFCN	—	Міжнародна мережа фактчекінгу (International Fact-Checking Network)
ІПСО	—	Інформаційно-психологічна операція
API	—	Інтерфейс прикладного програмування (Application Programming Interface)

NER	—	Розпізнавання іменованих сутностей (Named Entity Recognition)
ASR	—	Автоматичне розпізнавання мови (Automatic Speech Recognition)
OCR	—	Оптичне розпізнавання символів (Optical Character Recognition)
LDA	—	Латентне розміщення Діріхле (Latent Dirichlet Allocation)
NMF	—	Невід'ємна матрична факторизація (Non-negative Matrix Factorization)
CIB	—	Скоординована неавтентична поведінка (Coordinated Inauthentic Behavior)
SNA	—	Аналіз соціальних мереж (Social Network Analysis)
OSINT	—	Розвідка на основі відкритих джерел (Open Source Intelligence)
TTPs	—	Тактики, техніки та процедури (Tactics, Techniques, and Procedures)
HITL	—	Людина-в-циклі (Human-in-the-Loop)
MLOps	—	Операції машинного навчання (Machine Learning Operations)
MVP	—	Мінімально життєздатний продукт (Minimum Viable Product)
СОП	—	Стандартні операційні процедури
R&D	—	Дослідження та розробки (Research and Development)
UI	—	Інтерфейс користувача (User Interface)
UX	—	Досвід користувача (User Experience)

ВСТУП

В епоху глобалізації та стрімкого розвитку цифрових технологій інформаційна війна перетворилася на невід'ємний та потужний інструмент впливу в сучасних конфліктах та геополітичному протистоянні. Інформація та контроль над нею стали ключовими ресурсами, здатними формувати громадську думку, впливати на політичні процеси та дестабілізувати цілі суспільства. Особливого значення набувають соціальні мережі та онлайн-платформи, які, з одного боку, демократизували доступ до інформації, а з іншого – стали ефективним каналом для швидкого та масового поширення деструктивних наративів, дезінформації та пропаганди.

Сучасні інформаційні операції характеризуються високим рівнем координації, використанням передових технологій, таких як штучний інтелект, автоматизовані бот-мережі та методи психологічного впливу. Цілеспрямовані інформаційні атаки, що включають поширення фейкових новин, маніпулятивного контенту та пропагандистських меседжів, становлять серйозну загрозу національній безпеці, підриваючи довіру до державних інституцій, розпалюючи суспільні конфлікти та послаблюючи стійкість держави перед зовнішніми та внутрішніми викликами.

Кваліфікаційна робота присвячена дослідженню загроз інформаційної війни та розробці концептуальної моделі автоматизованого механізму протидії таким загрозам. У рамках роботи розглядаються теоретичні основи інформаційного протиборства, аналізуються сучасні методи та технології поширення деструктивного впливу, досліджується національний та міжнародний досвід протидії інформаційним атакам, а також проектується модульна архітектура запропонованої системи та здійснюється практична реалізація її ключових компонентів.

Актуальність роботи обумовлена безпрецедентною інтенсивністю та масштабами інформаційної агресії, з якою стикається Україна, особливо в

умовах повномасштабної війни з російською федерацією. Існуючі методи протидії часто є фрагментарними та недостатньо ефективними для оперативного реагування на динамічні та технологічно складні інформаційні атаки. Це зумовлює нагальну потребу в розробці та впровадженні комплексних, переважно автоматизованих систем моніторингу, аналізу та нейтралізації інформаційних загроз.

Метою роботи підвищення ефективності протидії інформаційним загрозам шляхом дослідження існуючих методів та розробки концептуальної моделі автоматизованого механізму, спрямованого на оптимізацію процесів їх виявлення, аналізу та реагування.

Об'єктом дослідження є процеси селекції і обробки інфромації під час інформаційної війни

Предметом дослідження є автоматизовані методи виявлення та протидії дезінформації, зокрема механізми аналізу, перевірки та маркування недостовірного контенту в медійному просторі.

Для досягнення зазначеної мети поставлено наступні завдання:

1) Проаналізувати теоретико-правові засади та сучасний стан проблеми інформаційної війни, включаючи класифікацію її загроз, роль соціальних мереж як інструменту впливу та особливості нормативного регулювання сфери інформаційної безпеки в Україні.

2) Дослідити існуючі методи та практичний досвід протидії інформаційним загрозам, проаналізувавши національні та міжнародні підходи, механізми виявлення дезінформації в онлайн-середовищі та оцінивши ефективність поточних рішень.

3) Розробити концептуальну модель автоматизованого механізму протидії інформаційним загрозам, обґрунтувавши його необхідність, визначивши ключові вимоги, архітектуру, функціональні компоненти та перспективи розвитку й інтеграції.

Структура роботи включає вступ, три основні розділи, висновки та список використаних джерел. Перший розділ присвячений аналізу загроз в умовах

інформаційної війни. У ньому розглядається поняття інформаційної війни і класифікація загроз, аналізується вплив соціальних мереж як ключового інструменту в інформаційній протидії та маніпуляції свідомістю, а також досліджується нормативно-правове регулювання сфери інформаційної безпеки в Україні.

Другий розділ зосереджений на дослідженні сучасних методів виявлення інформаційних атак та огляді практичних рішень у сфері інформаційної безпеки. Розглядається національний досвід України та міжнародний досвід у протидії інформаційній війні, аналізуються механізми виявлення фейкових повідомлень у соціальних мережах та досліджується ефективність існуючих рішень.

Третій розділ присвячений побудові моделі автоматизованої системи протидії інформаційним загрозам. У ньому визначаються вимоги до автоматизованого механізму, розглядаються теоретичні засади функціонування модулів, проектується архітектура та ключові компоненти механізму, описується практична реалізація та тестування компонента механізму, а також визначаються перспективи його розвитку та інтеграції.

У заключній частині роботи наведено загальні висновки та підсумовано результати дослідження, спрямовані на створення науково-методичного підґрунтя для розробки та впровадження ефективного механізму автоматизованої протидії інформаційній війні. Таким чином, дана кваліфікаційна робота є комплексним дослідженням, яке охоплює ключові аспекти протидії сучасним інформаційним загрозам та пропонує науково обґрунтовані підходи до їх нейтралізації в умовах гібридних конфліктів.

РОЗДІЛ 1
АНАЛІЗ ЗАГРОЗ В УМОВАХ ІНФОРМАЦІЙНОЇ ВІЙНИ

1.1. Поняття інформаційної війни і класифікація загроз

Формування концепції інформаційної війни має глибокі історичні корені, які сягають періоду до появи цифрових технологій. Упродовж XX століття, особливо в роки світових війн і під час холодної війни, інформація стала об’єктом маніпуляцій, а пропагандистські кампанії — ефективним інструментом впливу на свідомість мас. Саме в цей період спостерігалося зростання значення психологічного впливу, використання радіо та друкованих видань для трансляції вигідних меседжів, формування наративів та дискредитації супротивника. Проте, поняття «інформаційна війна» в його сучасному розумінні почало системно застосовуватись лише в 1990-х роках, зокрема після ухвалення у 1992 році директиви Міністерства оборони США DOD S 3600.1, в якій термін уперше з’явився на офіційному рівні. У подальшому цей підхід був розширений у звітах аналітичного центру RAND, що наголошували на можливості досягнення стратегічних переваг без застосування фізичної сили — виключно через маніпуляції інформаційним простором.

У період становлення постіндустріального суспільства інформація поступово перетворювалася на ключовий ресурс, що значно впливає на хід політичних, економічних і військових процесів. Відповідно, виникла необхідність перегляду традиційних уявлень про безпеку. До класичних підходів протистояння додалася стратегія інформаційного домінування, суть якої полягає у здобутті переваги шляхом контролю над інформаційними потоками та технологіями. У звітах RAND звертається увага на важливість взаємодії державного та приватного секторів у справі захисту інформаційної інфраструктури. Окремо акцентується на тому, що конфлікти нового типу можуть мати реальні наслідки, хоча й не супроводжуються прямим збройним

зіткненням. Таким чином, інформаційна війна розглядається не як допоміжний елемент, а як повноцінна складова сучасних конфліктів, що може визначати його результат.

У сучасному науковому дискурсі інформаційна війна розуміється як цілеспрямований вплив на свідомість, емоції та поведінку людей з метою досягнення політичних, військових або соціальних результатів. Цей вплив може здійснюватися через різноманітні канали — від традиційних медіа до соціальних мереж — і мати різні форми: від нав’язування певних інтерпретацій подій до поширення відверто неправдивих відомостей. Особливо небезпечними є сучасні технологій, такі як штучний інтелект, deepfake або автоматизовані бот-мережі, які дозволяють створювати надзвичайно реалістичний фальсифікований контент, автоматизувати його поширення у великих масштабах та значно ускладнювати його виявлення і спростування, тим самим підвищуючи ефективність та руйнівний потенціал інформаційних операцій.. Водночас важливим чинником успішності інформаційної атаки є не лише зміст повідомлення, а й форма подання, емоційне забарвлення та контекст, у якому ця інформація сприймається.

Таблиця 1.1

Класифікація загроз

джерело	державні суб’єкти
	недержавні суб’єкти
	індивідуальні суб'єкти
природа	дезінформація
	мізінформацій
	пропаганда
	кібератака
	психологічна маніпуляція
	операція впливу

продовження таблиці 1.1

мета	вплив на громадську думку
	ураження критичної інфраструктури
	втручання у політичні процеси
	деморалізацію збройних сил
	підрив міжнародних відносин
об’єкт впливу	когнітивний
	інформаціний
	фізичний

Для більш глибокого розуміння сутності інформаційної війни доцільно розглянути класифікацію її загроз за різними критеріями (таблиця 1.1). Одним з ключових підходів є поділ загроз за джерелом. Тут виокремлюють державні суб’єкти, такі як урядові структури та спецслужби, що володіють значними ресурсами для проведення масштабних інформаційних операцій. Та еедержавні суб’єкти, включаючи терористичні та екстремістські організації, хакерські групи та організовану злочинність, також становлять значну загрозу, часто керуючись ідеологічними або фінансовими мотивами. Окрему категорію складають індивідуальні суб’єкти, здатні завдавати шкоди як свідомо, так і несвідомо поширюючи дезінформацію.

Іншим важливим критерієм є природа загрози. До цієї категорії належать дезінформація (навмисне поширення неправдивої інформації), мізінформація (ненавмисне поширення неточної інформації), пропаганда (систематичне поширення упередженої інформації), кібератаки (зловмисні дії в кіберпросторі), психологічна маніпуляція (вплив на емоції та поведінку) та операції впливу (скоординовані зусилля для формування громадської думки). Кожен з цих типів загроз має свої особливості та потребує специфічних методів протидії.

Крім того, загрози інформаційної війни можуть бути класифіковані за метою. Серед основних цілей виділяють вплив на громадську думку, ураження критичної інфраструктури, втручання у політичні процеси, деморалізацію

збройних сил та підрив міжнародних відносин. Розуміння мети атаки дозволяє прогнозувати її потенційні наслідки та розробляти правильні заходи реагування.

Нарешті, загрози розрізняються за об’єктом впливу, який може бути когнітивним (вплив на мислення), інформаційним (маніпулювання даними) або фізичним (реальні наслідки в матеріальному світі).

Таким чином, інформаційна війна є складним, багатокомпонентним явищем, яке вимагає системного дослідження, комплексного підходу та інтеграції зусиль як наукової спільноти, так і практиків у сфері національної безпеки. Актуальність цієї теми зростає в умовах гібридних конфліктів, зокрема на прикладі війни росії проти України, де інформаційний компонент має стратегічне значення. Наступні розділи роботи будуть присвячені поглибленому аналізу методів виявлення таких загроз, оцінці ефективності сучасних підходів та формуванню концепції автоматизованого механізму протидії.

1.2. Вплив соціальних мереж як ключового інструменту в інформаційній протидії та маніпуляції свідомістю

Продовжуючи розгляд еволюції засобів інформаційного впливу, варто зупинитися на тому, як саме трансформувалася роль каналів поширення інформації, починаючи з традиційних медіа і до появи цифрових платформ. Упродовж більшої частини XX століття засобами, через які здійснювався інформаційний вплив, залишалася переважно друковані газети, радіо та телебачення. Саме вони стали першими інструментами, що використовувались державами для ведення цілеспрямованої інформаційної боротьби, яка згодом набула системного характеру й трансформувалася в інформаційну війну.

У міжвоєнний період уряди низки країн почали усвідомлювати силу централізованого інформаційного контролю. Яскравим прикладом є використання радіо нацистською Німеччиною, де пропаганда була інституційно організованою через Міністерство народної просвіти та пропаганди. Подібні

механізми активно розвивалися в СРСР, де цензура і монополія на ЗМІ дозволяли державі формувати одностороннє уявлення про реальність.

Переломним моментом став початок масового доступу до інтернету в середині 1990-х років, що порушив централізовану структуру комунікації. Якщо раніше інформаційні потоки контролювалися переважно державами або великими корпораціями, то нові цифрові платформи відкрили можливість практично будь-кому впливати на масову свідомість. Поява блогів, форумів, а згодом і соціальних мереж означала перехід від вертикального до горизонтального поширення інформації. Цей зсув, що відбувся впродовж першого десятиліття XXI століття, змінив не лише способи комунікації, але й ускладнив виявлення джерел інформаційного впливу, що стало новим викликом у сфері безпеки.

Центральну роль у цьому процесі відіграють алгоритми, що управляють подачею контенту. Системи рекомендацій, на основі яких формується стрічка новин у таких платформах, як Facebook, TikTok чи YouTube, оптимізовані не на правдивість, а на залученість користувача. Внаслідок цього поширення отримують переважно ті матеріали, які викликають емоційну реакцію, а не ті, що відповідають критеріям достовірності. Таким чином, формується інформаційне середовище, у якому кожен користувач взаємодіє з обмеженим колом тем і поглядів, що підтримують його власні переконання. Це сприяє виникненню явищ «інформаційної бульбашки» та «резонансної камери», в яких альтернативні (часто правдиві) точки зору просто не отримують уваги.

Суттєвим фактором у поширенні дезінформації виступає сама поведінка користувачів. Пересічний учасник соціальної мережі може несвідомо поширювати неправдивий чи маніпулятивний контент, вважаючи його правдивим або емоційно значущим. Часто спрацьовує логіка довіри до джерел, що є "своїми" за ідеологічною або культурною ознакою, а популярність публікації починає сприйматись як свідчення її достовірності. У цьому контексті користувач виступає одночасно і як жертва, і як ретранслятор інформаційного впливу [1].

Особливої актуальності проблема набуває в умовах збройного конфлікту. Соціальні мережі в таких випадках перетворюються на поле бою, де інформація стає зброєю. У війні росії проти України ці платформи використовуються як для інформування населення, так і для поширення ворожих деструктивних наративів [2]. Наприклад, Telegram став водночас джерелом оперативної інформації та інструментом маніпуляцій, що ускладнює завдання розрізнення достовірного повідомлення від ворожого «вкиду».

Таким чином, соціальні мережі мають двояку природу. Вони є як простором вільного доступу до інформації, так і середовищем, у якому дезінформація може не лише існувати, а й активно поширюватися. Це зумовлює потребу в формуванні культури інформаційної гігієни, підвищенні цифрової грамотності населення та впровадженні нових механізмів протидії за інформаційному впливу. У подальшому дослідженні буде розглянуто ефективність сучасних стратегій та технологічних засобів, що використовуються для нейтралізації інформаційних загроз в соціальних мережах.

1.3. Нормативно-правове регулювання сфери інформаційної безпеки в Україні

Російсько-українська війна, що триває з 2014 року та особливо загострилася після повномасштабного вторгнення у 2022 році, виявила критичну потребу у формуванні системного підходу до захисту інформаційного простору України. Агресія російської федерації носить не лише військовий, а й інформаційний характер, де пропаганда, дезінформація та маніпуляції масовою свідомістю стали ключовими інструментами впливу. У відповідь Україна активізувала нормативно-правові механізми протидії інформаційним загрозам, зосереджуючи зусилля на зміцненні національної безпеки в медіасфері, культурі та публічному просторі. Основні положення ключових актів представлено в таблиці 1.2.

Таблиця 1.2

Законодавча база

Нормативно-правовий акт	**Короткий зміст і значення**
Закон України «Про засудження та заборону пропаганди російської імперської політики в Україні і деколонізацію топонімії» (2023)	Засуджує імперську політику РФ як злочинну, передбачає демонтаж символіки, перейменування топонімів, заборону «русского міра» в публічному просторі.
Закон України «Про медіа» (2023)	Впроваджує європейські стандарти, регулює всі типи медіа, забороняє вплив держави-агресора, передбачає спеціальні умови діяльності ЗМІ в умовах війни.
Закон України «Про кінематографію» (з оновленнями після 2014 р.)	Забороняє до показу фільми, що просувають російську пропаганду або героїзують спецслужби й військових РФ.
Закон України «Про основні засади забезпечення кібербезпеки України» (2017)	Формує правову базу для боротьби з кіберзагрозами, у т.ч. дезінформаційними атаками, передбачає взаємодію держави й приватного сектору.
Закон України «Про національну безпеку України» (2018)	Інформаційна безпека визначена складовою національної безпеки; передбачає заходи протидії маніпулятивному впливу в медіа.

Закон України «Про засудження та заборону пропаганди російської імперської політики в Україні і деколонізацію топонімії» став важливим інструментом деколонізаційної політики, спрямованої на системне очищення публічного простору України від символів та ідеологем, пов’язаних із російською імперською традицією. Визнання російської імперської політики злочинною в правовому полі означає офіційну легітимацію історичних оцінок, які раніше існували переважно в експертному та громадському дискурсі.

Закон передбачає механізми демонтажу пам’ятників, перейменування вулиць та населених пунктів, ліквідації юридичних осіб і назв, пов’язаних із російською імперською політикою, що функціонували в українському просторі упродовж десятиліть. Це не лише адміністративні дії, а й символічна деокупація ідентичності – перехід до власного національного наративу. В умовах інформаційної війни, коли Росія активно апелює до "спільної історії", "єдиної культури" та "православної цивілізації", відмова від імперських маркерів набуває стратегічного значення: вона підриває основи ідеологічного впливу та делегітимізує аргументи ворога на міжнародній арені.

Крім того, встановлення відповідальності за порушення заборони пропаганди посилює юридичну вагу цього акту, перетворюючи деколонізацію з декларативної ініціативи на обов’язкову до виконання державну політику. Застосування цього закону закладає основи для формування стійкого інформаційного імунітету.

Закон України «Про медіа». З ухваленням нового Закону «Про медіа» Україна здійснила довгоочікувану реформу в галузі медійного регулювання, яка відповідає європейським стандартам [3]. Закон консолідує правове регулювання телевізійного, радіо-, друкованого, онлайн- та інших видів медіа, зосереджуючи увагу на питаннях прозорості власності, недопущення впливу держави-агресора та захисту національного інформаційного суверенітету [1].

Особливої уваги заслуговує Розділ IX, що врегульовує діяльність медіа в умовах збройної агресії. Це вперше, коли у законі передбачено спеціальні норми, адаптовані до реалій війни: обмеження на поширення інформації, що виправдовує агресію чи дискредитує обороноздатність України; заборона власності з боку громадян або суб’єктів російської федерації; обмеження на діяльність іноземних медіа, що можуть використовуватися як інструмент інформаційного тиску.

Важливим є розширення повноважень Національної ради України з питань телебачення і радіомовлення. Цей орган отримав більше інструментів для оперативного реагування на інформаційні загрози, включаючи моніторинг

онлайн-контенту, що раніше залишався поза прямою юрисдикцією. Встановлення обов'язкових стандартів об'єктивності, перевірки фактів і недопущення мови ворожнечі створює правове середовище, що сприяє формуванню відповідального медіа-середовища.

Закон України «Про кінематографію». Хоча кінематограф може видаватися віддаленим від сфери безпеки, саме через культурну продукцію часто здійснюється м'яка пропаганда. Закон «Про кінематографію» регламентує механізми допуску фільмів до прокату, можливості державної підтримки кіновиробництва, а також визначає критерії, за якими фільм може бути заборонений до демонстрації.

Зміни, внесені до закону після 2014 року, посилили механізми блокування ворожої пропаганди, зокрема заборону фільмів, у яких прославляються представники збройних формувань РФ, співробітники спецслужб, або ж популяризуються ідеї "русского міра". Таким чином, кінематографічна політика стала частиною загального комплексу заходів із протидії гібридному впливу.

У воєнний час культурна безпека набуває прикладного значення: демонстрація чи доступність фільмів, що містять елементи ворожої пропаганди, може бути не менш небезпечною за інформаційні зведення. Закон забезпечує можливість превентивного контролю за медіапродуктом, що стає бар'єром для проникнення пропагандистських наративів.

Базовими рамковими законами, що створюють фундамент для політики інформаційної безпеки є зокрема, Закон України «Про основні засади забезпечення кібербезпеки України» (2017) визначає організаційно-правові механізми протидії кіберзагрозам [4], включно з дезінформаційними атаками, та передбачає взаємодію державних органів, суб'єктів господарювання і громадянського суспільства. У свою чергу, Закон України «Про національну безпеку України» (2018) закріплює інформаційну безпеку як невід'ємний елемент загальнонаціональної безпеки [5], а також зобов'язує державу вживати заходів для запобігання маніпулятивному впливу в медіа, що посилює нормативну основу для дій у межах гібридної війни.

Висновки до розділу 1

У першому розділі кваліфікаційної роботи було проаналізовано теоретико-правові засади та сучасний стан проблеми інформаційної війни, включаючи класифікацію її загроз, роль соціальних мереж як інструменту впливу та особливості нормативного регулювання сфери інформаційної безпеки в Україні.

У ході аналізу було встановлено, що інформаційна війна є складною, багатоаспектною діяльністю, спрямованою на досягнення стратегічних переваг шляхом маніпулювання інформацією, впливу на суспільну свідомість, цінності та поведінку цільових аудиторій. Розглянуто еволюцію підходів до розуміння цього явища та його ключові характеристики в сучасному глобалізованому світі.

Було проведено класифікацію основних загроз, що виникають в умовах інформаційної війни. До них належать поширення дезінформації та фейкових новин, цілеспрямовані пропагандистські кампанії, операції впливу (ІПСО), кібератаки на інформаційну інфраструктуру, а також використання маніпулятивних технологій, таких як діпфейки та скоординовані мережі ботів.

Також розглянуто особливості нормативно-правового регулювання сфери інформаційної безпеки в Україні. Проаналізовано ключові законодавчі акти, стратегії та доктрини, що визначають державну політику у цій сфері. Виявлено як досягнення, так і наявні прогалини та виклики, зокрема щодо регулювання новітніх онлайн-платформ та месенджерів, а також щодо забезпечення ефективної координації зусиль державних органів.

Результати, отримані в першому розділі - розуміння природи інформаційних загроз, каналів їх поширення та стану нормативно-правового забезпечення, створюють необхідне теоретичне підґрунтя для подальшого дослідження існуючих методів протидії та розробки ефективного механізму автоматизованого реагування.

РОЗДІЛ 2
ДОСЛІДЖЕННЯ СУЧАСНИХ МЕТОДІВ ТА ПРАКТИЧНИХ РІШЕНЬ У ПРОТИДІЇ ІНФОРМАЦІЙНІЙ ВІЙНІ.

2.1. Національний досвід України у виявленні та нейтралізації інформаційних загроз

В стані інформаційної війни з Росією: з початком широкомасштабного вторгнення Кремль розгорнув потужну кампанію пропаганди в медіа. Уже в перші дні повномасштабного вторгнення російські «інформаційні війська» створили більше сотні Telegram-каналів для різних українських міст, під видом локальних новин, які насправді поширюювали проросійські наративи (рисунок 2.1).

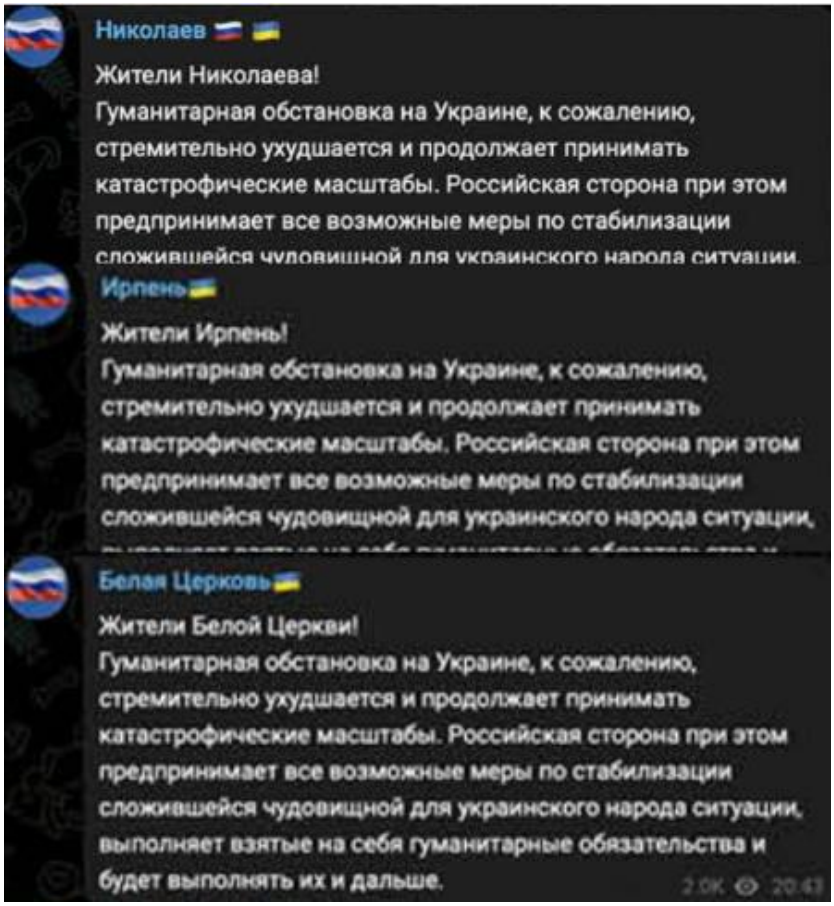


Рисунок 2.1 - «Телеграм окупація» мережею російських ботів

Це стало одним із прикладів наміреної інформаційної атаки. Пізніше український сегмент Telegram стрімко поповнювався анонімними каналами, які створювали видимість поразки в військовому просторі. Так, дані дослідження «Детектора медіа» свідчать про понад 400 таких проросійських/окупаційних каналів [6]. При цьому Telegram став головним джерелом новин для більшості українців після початку війни: опитування Internews показало, що 60 % респондентів надають перевагу Telegram як основному інформаційному каналу. Ця ситуація підсилює потребу в ефективних державних механізмах виявлення та нейтралізації інформаційних загроз [7].

Державні органи протидії дезінформації в Україні (таблиця 2.1) формувалися переважно після 2021 року. Зокрема, при РНБО було створено Центр протидії дезінформації (ЦПД), покликаний моніторити інформаційний простір і координувати реагування [8], також під наглядом держави залишилися Нацрада з питань телебачення і радіомовлення (що регулює традиційні медіа) та СБУ (що відповідає за контррозвідувальні заходи в кіберпросторі).

Таблиця 2.1

Державні органи протидії дезінформації в Україні

Орган/Організація	Основні функції протидії пропаганді	Сфера впливу/Об'єкти	Виклики/Обмеження
Центр протидії дезінформації (ЦПД) при РНБО	Моніторинг, аналіз, координація реагування	Загальний інформаційний простір, онлайн-медіа, соцмережі	Координаційні функції, потреба в посиленні повноважень
Служба безпеки України (СБУ)	Контррозвідувальні заходи, виявлення ІПСО, блокування	Кіберпростір, ворожі ресурси	Реактивний характер, складність взаємодії

продовження таблиці 2.1

Національна рада з питань телебачення і радіомовлення	Регулювання діяльності традиційних та онлайн-ЗМІ (зареєстрованих)	Традиційні ЗМІ, зареєстровані онлайн-медіа	Обмежений вплив на месенджери та анонімні ресурси
Недержавні фактчекінгові організації (StopFake, VoxCheck тощо)	Верифікація інформації, спростування фейків, навчання	Публічний інформаційний простір, соцмережі	Залежність від фінансування, обмежені ресурси

Проте їхня діяльність стикається з низкою проблем. По-перше, законодавство і процедура реагування є доволі повільними та формалізованими. Зокрема, Закон «Про медіа» взагалі не поширюється на месенджери: представник Нацради Олександр Бурмагін визнав, що «закон жодним чином не регулює Telegram», що є «головним недоліком» нинішньої системи регулювання. Водночас для державних службовців було пропоновано заборонити користування Telegram на робочих пристроях, що унеможливлює їхнє оперативне занурення у цей інформаційний канал. По-друге, робота органів розпорошена та фрагментована: не було єдиного централізованого механізму координації. Лише у 2025 році Нацрада ініціювала створення міжвідомчої робочої групи для консолідації зусиль різних відомств у боротьбі з дезінформацією. Та поки що це об’єднання не охопило всіх залучених інституцій, і кожна структура працює частково на своєму напрямку. По-третє, існуючі державні рішення мають обмежений вплив на масовий аудиторійний простір, особливо в онлайн-медіа і месенджерах. Наприклад, СБУ було повідомлено, що з початку війни передала в Telegram скарги на понад 1500 пропагандистських каналів, але ці дії фактично стосуються лише окремих випадків (канали часто змінюють назви та хости, лишаючись поза досяжністю).

У той же час Нацрада може впливати лиш на зареєстровані ЗМІ — однак на багатьох популярних інформаційних ресурсах (сайтах і каналах у соцмережах) часто відсутня чітка реєстрація, або робота медіаресурсу й зовсім відбувається анонімно, не підпорядковуючись жодному українському регулятору. Отже, через технічні і законодавчі обмеження державні механізми не спроможні повністю охопити ключові інформаційні ресурси.

У цьому контексті важливу роль відіграють недержавні ініціативи та незалежні медіааналітики. Протягом війни утворилися численні громадські та експертні проєкти з фактчекінгу та спростування фейків – від великих організацій (напр. StopFake, VoxCheck) до волонтерських телеграм-каналів і блогів [9]. Ці учасники інформаційного простору оперативно реагують на маніпуляції, часто першим надаючи альтернативні дані і пояснення. Однак і тут є суттєві труднощі. Значна частина таких ініціатив залежить від популярності контенту та залучення аудиторії: у прагненні до переглядів вони можуть публікувати спрощені або клікбейтні матеріали, акцентуючись на збільшенні охоплення. В своєю чергу, проєкти, які прагнуть чесного аналізу, часто працюють на невеликі грантові бюджети і у разі скорочення фінансування ризикують збанкрутувати. Так, незалежна платформа «Детектор медіа» виявила, що після тимчасового замороження фінансової допомоги США українських фактчекерів їхня присутність у цитуваннях медіа суттєво знизилася – місце на інфополі зайняли передусім міжнародні проєкти та державні комунікаційні ініціативи. Про серйозну загрозу діяльності незалежних ЗМІ і фактчекерів через заморожування американської допомоги попереджала і керівниця інформаційної безпеки РНБО Н. Ткачук. Таким чином, хоча недержавні організації відіграють критичну роль у верифікації інформації, їхня робота нестійка: без постійного фінансового ресурсу навіть найякісніші проєкти ризикують зникнути.

Узагальнюючи, можна констатувати, що в Україні сформовано низку ініціатив для протидії інформаційним загрозам, але наявні механізми далекі від повноцінної ефективності. З одного боку, держструктури через бюрократичні та законодавчі обмеження запізнюються з реакцією і не здатні працювати

масштабно на нових платформах, особливо в Telegram. З іншого – недержавні фактчекери хоча й демонструють оперативність, однак їх часто не вистачає ресурсів і мотивації забезпечувати безперервну роботу. З огляду на це можна зробити висновок: існуючих заходів замало для протидії агресивній інформаційній війні, і необхідно вживати додаткові інструменти та рішення.

2.2. Міжнародний досвід у протидії інформаційній війні

НАТО (англ. North Atlantic Treaty Organization — NATO) — північноатлантичний альянс із моменту свого створення в 1949 році зазнає ворожі інформаційні напади. Спершу з боку радянської пропагандиської машини, а потім росії. Як наслідок після початку російсько-української війни, що розпочалася з окупації Криму у 2014 році, організація активно протидіє зростанню дезінформації навколо власної діяльності.

Інформаційні загрози у вигляді цілеспрямованої, скоординованої діяльності, а саме маніпуляція даними, поширення неправдивої інформації через мультимедіа, спрямовані на різні верстви населення — як на тих, хто стежить за традиційними медіа, так і на тих, хто віддає перевагу соціальним мережам. Вони створюються з метою дестабілізації суспільства через поглиблення суперечностей і плутанини, підриву довіри до державних установ та, зрештою, ослаблення Альянсу.

Тож перед НАТО стоїть виклик протистояти цим загрозам і забезпечити стійкості за допомогою різних механізмів, у тому числі співпраці з державами-членами[10].

Оскільки загроза ослаблення Північноатлантичного союзу стосується кожну країну-членкиню, стратегія НАТО у сфері боротьби з інформаційними загрозами спирається на тісну співпрацю зі залученими державами. Такий підхід створює повну картину поточних загроз для кожного учасника. До того ж, це сприяє розвитку швидкого виявлення та реагування на зловмисні дії.

Підхід співпраці породжує різнорівневі реагувальні заходи, а саме коротко-, середньо- та довгострокові, зокрема, й випереджувальні. Вони дають змогу ефективно адаптуватися до ситуації та коригувати масштаби зусиль відповідно до рівня загрози. Для досягнення стратегічних комунікаційних цілей НАТО базує свої дії на чотирьох ключових напрямах: розуміння, запобігання, стримування і пом'якшення наслідків та відновлення.

1) Розуміння. Перед Альянсом стоїть потреба в оцінці інформаційного середовища, аби мати змогу чітко та якісно реагувати. Для цього команда спеціалістів займається моніторингом і аналізом інформації, що стосується Альянсу. Це сприяє розумінню поточних загроз, виявленню джерела ворожих наративів і визначення способів подальшої протидії[10].

2) Запобігання. Наріжним каменем у запобіганні розпаду союзу внаслідок інформаційних атак, які підривають довіру громадян до демократичного альянсу, є прозора й відкрита комунікація. Передбачаючи ворожі наративи та випереджаючи їхнє поширення у рамках Групи швидкого реагування НАТО, союз зміцнює стійкість суспільства до дезінформації, тим самим зменшуючи дієвість загроз. Для комунікації з населенням Альянс залучає різноманітні медіа, від газет до соціальних мереж, аби донести потрібну інформацію якомога більшій аудиторії.

3) Стримування і пом'якшення наслідків. Доволі часто ситуація складається так, що провокаційні новини поширюються через значну кількість фальшивих акаунтів, які заповнюють інформармаційний простір дезінформацією, спрямованою на виклик бурхливої емоційної реакції у читача[10]. Наприклад, компанія Twitter заявила у 2018 році про видалення десятків мільйонів облікових записів через наростання проблеми продажу фальшивих підписників медійним особам та політикам, зокрема. Це був важливий крок для власників соціальної мереж, адже їхній продукт був важливим інструментом впливу росії під час перемоги Трампа у 2016 році. росія у своїх цілях використовувала десятки тисяч фейкових акаунтів для поширення вигідної їй пропаганди[11]. Тож, повертаючись до головної теми, для НАТО у

свою чергу найкращим способом боротьби з подібними загрозами є визнання їхньої наявності та уникнення реакції на них. Проте трапляються ситуації, коли певні тези набули переконливих розмірів, тому їхній вплив на суспільство неможливо не помітити. У такому разі Альянс змушений публічно спростовувати наклепи у свій бік.

4) Відновлення. Кваліфікований штат людей проводить вичерпний аналіз здійсненої інформаційної атаки. Оцінює тактики, застосовувані зловмисниками, та досліджує вразливості інформаційно-безпекової структури, якими користуються оператори інформаційного впливу. Такі дії допомагають Альянсу сформувати більш ефективну стратегію реагування та підвищити стійкість суспільства до майбутніх атак[10].

Отже, стратегія НАТО у протидії дезінформації базується на чотирьох пунктах: розуміння(1), запобігання(2), стримування і пом'якшення наслідків(3), відновлення(4). Оскільки перший і четвертий пункти стосуються розуміння дій ворожих сил, а другий та третій – комунікації з населенням, то автор вважає доцільним описати чіткі впровадження введені Північноатлантичним альянсом, спираючись на такі основні цілі протистояння, як розуміння та взаємодія.

Розуміння. Досліджуючи дану тематику, неможливо не згадати про Центр стратегічних комунікацій НАТО (англ. NATO Strategic Communications Centre of Excellence — NATO StratCom COE), який розпочав роботу в січні 2014 року та фінансується внесками зацікавлених країн-спонсорів.[12] Дослідження у сфері інформаційної безпеки базується на поєднанні знань із військового, приватного, державного та академічного секторів. Для цього сформовано команду експертів із 17 країн, яка займається розпізнаванням і управлінням загрозами.

До основних цілей роботи входять стратегічні комунікації, тенденції у цифровій безпеці та розуміння стратегій ворожих атак. На даному етапі дослідження, цікавим є розглянути одну з найвизначніших подій – реалізація платформи «InfoRange» в рамках щорічних навчань «Locked Shields».

Платформа моделювання інформаційного середовища «InfoRange» розроблена, для динамічного навчання фахівців у галузі стратегічної комунікації,

включно з публічною дипломатією, цивільними і військовими зв'язками з громадськістю, інформаційними та психологічними операціями. Платформа дозволяє моделювати реальні інформаційні середовища, відтворюючи мережі, взаємодії та процеси. Таким чином створюється комплексне симуляційне середовище, у якому фахівці отримують можливість здійснювати спостереження та аналізувати роль стратегічних комунікацій у кризових ситуаціях, відповідно до власної спеціалізації. [13] Розробка таких технологій значною мірою сприяє якісній підготовці в умовах сучасних інформаційних війн. Розробники підкреслюють, що стійкість проти зловмисних атак вимагає навчання, яке відповідає сучасним викликам. Оскільки генеративний штучний інтелект (ШІ) стає доступнішим, його необхідно інтегрувати у навчання StartCom, зокрема в процесах пов'язаних зі симуляціями середовищ.[14]

Для кращого ознайомлення з темою, необхідно також детальніше згадати про «Locked Shields» та їхніх організаторів. Це щорічні змагання, які організовує Об'єднаний центр передових технологій з кібероборони НАТО (англ. NATO Cooperative Cyber Defence Centre of Excellence – CCDCOE). На цьому заході змагається дві команди, беручи на себе роль національних кібергруп швидкого реагування в умовах симуляції. Їхньою метою стоїть допомога вигаданій країні впоратися з масштабним кіберінцидентом з усіма його наслідками.Команди мають ефективно реагувати на інциденти та вирішувати стратегічні, юридичні й медійні завдання. Locked Shields моделює сучасні загрози на основі реалістичних сценаріїв і передових технологій.[15]

Говорячи про організаторів, CCDCOE проводять та поширюють масу науково-технічних досліджень з питань кібербезпеки та інформаційних атак. Крім цього, вже кілька років поспіль CCDCOE підтримує студентський конкурс бакалаврату «Глобальний інноваційний конкурс Вільяма та Мері» (англ. the William & Mary's Global Innovation Challenge – WMGIC). Такий союз вилився в конкурс «WMGIC x NATO Countering Disinformation Challenge 2024», об'єднавши студентів бакалаврату, науковців та фахівців галузі інформаційної безпеки задля розробки інноваційних рішень нагальних проблем дезінформації.

Серед ряду питань, що піднімалися, мала місце і російсько-українська війна. У таблиці 2.2 наведені підсумки деяких відкритих досліджень на дану тему[16].

Таблиця 2.2

Підсумки деяких відкритих досліджень

Назва колективу	Резюме
Team 20	Для протидії російській дезінформації про війну в Україні Team 20 рекомендує **розширити присутність НАТО в (соціальних) медіа**. Вони хочуть зробити канали НАТО більш доступними, надаючи інформацію мовами країн-членів і враховуючи культурні контексти.
HAWKEYES	Hawkeyes **рекомендують вищим навчальним закладам країн-членів НАТО надати доступ до платформи Ground News**. Це агрегатор, який базується в Канаді і збирає інформацію з 50 000 основних новинних видань і джерел з усього світу. Ця інформація буде зібрана в одному центральному місці, і кожна стаття буде перевірятися на достовірність і фактологічність 3 різними незалежними моніторинговими організаціями.
INFO GUARDIAN	INFO GUARDIAN пропонує **створити гру, в якій будуть представлені 5 ключових порад НАТО щодо боротьби з дезінформацією**, інтегровані в персонажів та сюжетну лінію. На початку та під час ключових переходів гравець буде стикатися з дезінформацією, яку Росія використовує для легітимізації свого несправедливого вторгнення в Україну. Наприкінці гри гравці будуть більш поінформовані та краще підготовлені до розпізнавання та протидії дезінформації.

продовження таблиці 2.2

Aleph	Aleph рекомендує **створити Північноатлантичне антидезінформаційне агентство** (англ. North Atlantic Anti-Disinformation Agency – NAADA) для фінансування організацій громадянського суспільства, стандартизації керівних принципів щодо дезінформації та моніторингу впливу стратегій
GREEC-INFORMATION	GREECINFORMATION пропонує план дій «Смілива заява», спрямований на посилення спроможності НАТО підтримувати зусилля України щодо протидії дезінформації. План дій складається з двох інструментів: **розробка алгоритму виявлення ботів**, здатного виявляти, аналізувати і повідомляти про поведінку керованих Росією ботів, які поширюють дезінформацію в соціальних медіа. **І кібер-інфраструктура**, спрямована на посилення цифрової стійкості України, у співпраці з агентствами ЄС.

Взаємодія. Вона є ключовим чинником у формуванні суспільної стійкості проти інформаційних атак. Альянс дотримується таких принципів, як відкритість, прозорість та доступність комунікацій. Задля цього застосовуються різноманітні канали впливу: соціальні мережі, зв'язки із ЗМІ, а також власна офіційна вебсторінка. Публікуючи деталі про військові навчання, наприклад, НАТО демонструє послідовну відкритість у комунікаціях та доносить широкій громадськості точну і своєчасну інформацію.

Окрім взаємодії з громадськістю, важливим залишається безпосередній контакт із державами-членкинями та партнерами. Створюються умови, в яких залучені сторони отримують змогу обмінюватися набутими знаннями і поширювати їх далі через свої громади. Таким чином Альянс формує такий собі ланцюг стійкості, що охоплює широкі маси.

З метою залучення молоді, проводяться кампанії на кшталт “Захисти майбутнє” [17], у межах якої молоді лідери думок та творці контенту отримують змогу дізнатися більше про діяльність Альянсу і поширювати ці знання у власному середовищі.

Варто також згадати і про додаткову підтримку неурядових організацій аналітичних центрів та університети через грантові програми. Діяльність спонсорованих організацій спрямована на формування проактивного, свідомого суспільства, здатного чиниити опір інформаційним впливам[18].

2.3. Механізми виявлення фейкових повідомлень у соціальних мережах

У сучасних умовах глобального інформаційного середовища, що характеризується високою швидкістю розповсюдження контенту, різноманітністю джерел та зростаючою кількістю інформаційних загроз, ефективне виявлення дезінформації набуває особливого значення для забезпечення інформаційної безпеки. У цьому контексті важливою є не лише наявність інструментів для боротьби з фейками, а й розуміння переваг, недоліків та доцільності застосування різних підходів залежно від конкретних умов.

Існуючі методи виявлення дезінформації можна умовно класифікувати на три основні групи: традиційні (з участю людини), автоматизовані (машинні) або гібридні. До традиційних належать фактчекінг, експертна оцінка та контекстуальний аналіз. Їх перевагою є глибоке розуміння змісту, культурного контексту та мовних особливостей повідомлення. Такі методи дозволяють оцінити не лише фактичну достовірність інформації, а й її емоційний вплив, потенціал до маніпулювання свідомістю аудиторії[19]. Водночас вони мають низку обмежень: високу трудомісткість, обмежену масштабованість і залежність від людського фактора, що унеможливлює оперативну обробку великого обсягу даних.

Автоматизовані методи базуються на використанні алгоритмів штучного інтелекту, машинного навчання, обробки природної мови (NLP), а також на аналізі метаданих (наприклад, джерела публікації, частоти ключових слів, мережевих зв’язків). Ці інструменти здатні здійснювати швидку перевірку величезних обсягів тексту, виявляти типові ознаки фейків, проводити класифікацію повідомлень за рівнем достовірності. Основною перевагою таких підходів є маштабність і швидкість, що дає змогу оперативно реагувати на інформаційні загрози. Проте машинні алгоритми часто не здатні точно враховувати контекст, іронію, сатиру чи культурно-обумовлені змісти, що призводить до помилкових оцінок. Крім того, деякі з них вимагають великих обсягів навчальних даних, які мають бути якісно класифіковані, що вже є трудозатратним завданням.

Гібридні моделі, які поєднують автоматизовану обробку з наступною перевіркою результатів людиною. Такий підхід дозволяє частково компенсувати недоліки кожного окремого методу, підвищуючи точність аналізу, зберігаючи при цьому достатню швидкість реагування. Наприклад, автоматичне виявлення підозрілих повідомлень може слугувати тригером для подальшої ручної перевірки фактчекерами.

Таким чином, ефективна протидія дезінформації має базуватись на комплексному використанні наявних інструментів із урахуванням їх функціональних можливостей, обмежень і контексту застосування. Підвищення результативності таких систем можливе шляхом удосконалення алгоритмів машинного навчання, розширення баз даних для тренування моделей, впровадження механізмів зворотного зв’язку з користувачами та підтримки співпраці між технологічними компаніями, державними структурами та громадянським суспільством.

Зважаючи на класифікацію методів виявлення дезінформації , їх переваги та недоліки, а також необхідність комплексного підходу для ефективної протидії інформаційним загрозам, важливо проаналізувати, як ці методи працюють на практиці. Враховуючи розглянуту важливість соцмереж, саме тут швидкість

розповсюдження контенту, обсяги даних та різноманіття форм подачі інформації створюють унікальні виклики та вимагають впровадження адаптованих механізмів протидії. Тому подальший аналіз буде зосереджений на розгляді конкретних стратегій та інструментів, що використовуються провідними соціальними платформами для виявлення та обмеження поширення дезінформації, а також оцінці ефективності.

Meta (Facebook, Instagram): незалежна перевірка фактів і обмеження дезінформації

Таблиця 2.3

Переваги та недоліки Meta	
Переваги	залучення незалежних експертів забезпечує додатковий контекст і сприяє зменшенню охоплення фейкового контенту
Недоліки	повільність людського фактчекінгу та обмежене охоплення
	нестабільність політики компанії – згортання ініціативи

Платформи Meta з 2016 року запровадили програму незалежного фактчекінгу, в межах якої публікації з підозрілим змістом передаються сертифікованим фактчекерам[20], що діють згідно з принципами Мережі незалежного фактчекінгу (IFCN). Перевірений контент позначається як «хибний», «частково хибний» тощо, після чого алгоритми Meta автоматично знижують його видимість або обмежують поширення. Крім того, компанія використовувала системи машинного навчання для виявлення змінених версій відомих фейків, наприклад, мемів або зображень із маніпулятивним текстом.

Meta активно співпрацювала з міжнародними фактчекерськими мережами, зокрема з IFCN/Poynter, підтримуючи регіональні ініціативи. Проте у 2025 році компанія оголосила про згортання програми в США та відмову від автоматичного обмеження контенту, який отримав фактчек-позначки, з переходом до моделі Community Notes[21]. Також діяльність Meta у росії було

обмежено, частково заблоковано доступ до Facebook після відмови припинити перевірку фактів щодо державних ЗМІ.

Переваги та недоліки такого методу вказані в таблиці Таблиця 2.3

X (Twitter): Community Notes, спільнотна модерація та боротьба з маніпуляціями

Таблиця 2.4

Переваги та недоліки X (Twitter)	
Переваги	децентралізація рішень і прозорість процесу включення користувачів у боротьбу з дезінформацією стимулює громадську відповідальність
Недоліки	низьке охоплення: багато користувачів не бачать приміток
	вразливість до координації ботів або заангажованих груп
	повільна реакція на маловідомі або довгі пости

Соціальна мережа X (раніше Twitter) запровадила систему Birdwatch (згодом перейменовану на Community Notes), яка дозволяє користувачам самостійно додавати контекст до публікацій. Примітки публікуються лише після схвалення користувачами — згідно з алгоритмом «bridging», що базується на досягненні міжгрупового консенсусу. Така модель модерації є децентралізованою і прозорою, адже адміністратори не втручаються напряму в процес маркування повідомлень.

Платформа також застосовує політику щодо «штучно змінених» медіа — вимагає маркувати або видаляти синтетичний або змінений контент[22], якщо він здатен завдати шкоди користувачам.

Переваги та недоліки такого методу вказані в таблиці Таблиця 2.4

YouTube: автоматизоване виявлення, зменшення охоплення та підтримка достовірного відеоконтенту

Таблиця 2.5

Переваги та недоліки YouTube	
Переваги	системний підхід до виявлення і обмеження поширення дезінформації
	використання масштабних обчислювальних ресурсів та незалежних джерел
Недоліки	відсутність прозорості щодо конкретних рішень
	повільне реагування на нові теми

YouTube реалізує стратегію «4R» (remove, reduce, raise, reward), що передбачає видалення порушень, обмеження охоплення сумнівного контенту[23], просування авторитетних джерел та підтримку надійних авторів. Сервіс забороняє відео з потенційно шкідливою дезінформацією, зокрема у сфері охорони здоров'я, а також конспірологічні та сфабриковані матеріали.

Контроль контенту здійснюється за допомогою поєднання ШІ-систем і ручної модерації, що дозволяє оперативно виявляти порушення. Алгоритми також не просувають «прикордонний» контент, тобто матеріали, що наближаються до порушень, але формально їх не перевищують. Додатково, у відео можуть з'являтися інформаційні панелі з перевіреною інформацією від організацій, сертифікованих IFCN.

Переваги та недоліки такого методу вказані в таблиці Таблиця 2.5

Telegram: децентралізований месенджер без алгоритмічної модерації

Таблиця 2.6

Переваги та недоліки Telegram	
Переваги	високий рівень приватності і свободи висловлювань
	можливість вільного обміну інформацією без централізованого втручання

продовження таблиці 2.6

Недоліки	відсутність модерації і механізмів перевірки сприяє масовому поширенню фейків
	особливо небезпечний у кризові періоди, коли інформаційне перевантаження ускладнює перевірку даних

На відміну від платформ із централізованим контролем, Telegram не застосовує алгоритмів для виявлення або обмеження поширення фейкового контенту. Уся відповідальність за модерацію покладається на адміністраторів окремих каналів або самих користувачів[24]. Повідомлення подаються у хронологічному порядку, що сприяє довготривалому перебуванню неправдивих повідомлень у стрічці.

Telegram особливо активно використовується в умовах російсько-української війни. Обидві сторони активно застосовують через високу анонімність і шифрування. Telegram залишається практично неконтрольованим навіть з боку державних структур.

Переваги та недоліки такого методу вказані в таблиці Таблиця 2.6

2.4. Дослідження ефективності існуючих рішень

У межах цього дослідження проаналізовано поширення фейкової, маніпулятивної та поляризуючої інформації на чотирьох основних платформах — Facebook, X/Twitter, Telegram та YouTube — у період після початку повномасштабного вторгнення росії в Україну (2022–2024 роки). Метою є виявлення загрозливих тенденцій у сфері інформаційної безпеки та визначення найуразливіших інформаційних середовищ з погляду дезінформації та пропаганди (таблиця 2.7).

Для цього було відібрано вибірку з найпопулярніших публікацій у кожному сегменті (за мовою та платформою), які отримали значне охоплення (кількість переглядів, поширень або взаємодій). Кожна публікація аналізувалася

за п’ятьма ключовими критеріями, і оцінкою в таблиці вважаємо відсоток публікацій із виявленими ознаками відповідного типу впливу серед цієї вибірки (тобто з *X* новин – скільки відсотків виявилися фейковими, маніпулятивними тощо). Для підвищення точності оцінювання було враховано мовну специфіку – українську, російську та англійську мови, оскільки мовна аудиторія може суттєво впливати на тип і зміст поширюваної інформації.

Неправдива інформація (фейк) – відомості, що не відповідають дійсності та спростовані офіційними джерелами, фактчекерами або незалежними ЗМІ.

Маніпулятивне подання – правдиві факти, подані із спотвореним контекстом, емоційним тиском або однобічною інтерпретацією з метою викривлення сприйняття.

Повторювані пропагандистські меседжі – наративи, що багаторазово відтворюються у різних публікаціях із характерною прокремлівською риторикою: дискредитація влади, просування зневіри в перемогу України, просування ворожих наративів.

Поляризуючий контент – матеріали, що мають на меті штучно розпалити внутрішні протистояння між групами населення, поширити ненависть, ворожнечу, агресію.

Анонімні публікації – контент, що не має чітко визначеного автора, видавця чи редакційної відповідальності, як правило, поширюється без перевірених джерел або контактної інформації, що ускладнює перевірку достовірності.

Таблиця 2.7

Наповненість соцмереж різними проявами інформаційної війни

Соцмережа	Мова	Неправдива інформація	Маніпулятивний контент	Повторювані наративи	Поляризуючий контент	Анонімність
Facebook	укр	34	43	26	29	22
	рос	47	56	37	44	27
	англ	27	39	24	31	17

продовження таблиці 2.7

X/Twitter	укр	36	47	33	41	28
	рос	49	67	46	57	36
	англ	31	54	39	42	26
Telegram	укр	53	74	63	73	89
	рос	64	83	76	84	86
	англ	29	47	36	49	62
YouTube	укр	27	37	23	32	12
	рос	39	54	29	47	16
	англ	19	29	17	26	7

Аналіз підтвердив, що Telegram є найуразливішою платформою для поширення дезінформації. У його україномовному сегменті 53% публікацій містили неправдивий контент, а 74% — маніпулятивне подання фактів. При цьому 63% повідомлень тиражували пропагандистські наративи, а 73% носили поляризуючий характер. Особливо високий показник анонімності: 89% каналів не мають ідентифікованого автора. У російськомовному Telegram ситуація ще гірша: 64% фейків, 83% маніпуляцій, 76% повторюваних меседжів, 84% поляризації та 86% анонімності. Навіть англомовні Telegram-канали показали значні проблеми (29% фейків, 47% маніпуляцій, 36% повторів, 49% поляризації, 62% анонімності).

Facebook демонструє значно кращі показники: в україномовному сегменті фейки становили 34%, маніпуляції — 43%, повторювані наративи — 26%, поляризація — 29%, а анонімних публікацій — 22%. Російськомовний Facebook показав дещо вищі цифри (47% фейків, 56% маніпуляцій, 37% повторів, 44% поляризації, 27% анонімності), тоді як англомовний сегмент залишався відносно чистим.

У X/Twitter україномовний контент містив 36% фейків, 47% маніпуляцій, 33% повторів, 41% поляризації і 28% анонімності. Російськомовний сегмент відзначився найвищими цифрами на цьому майданчику — до 67% маніпуляцій і

57% поляризації, а анонімні акаунти займали 36%. Англійський X/Twitter мав проміжні показники.

Нарешті, YouTube утримує найнижчі рівні деструктивного контенту: україномовний — 27% фейків, 37% маніпуляцій, 23% повторів, 32% поляризації, 12% анонімності. Російськомовний сегмент мав 39% фейків, 54% маніпуляцій, 29% повторів, 47% поляризації та 16% анонімності, тоді як англомовний показав найменші значення.

Висновки до розділу 2

У другому розділі кваліфікаційної роботи було досліджено існуючі методи та практичний досвід протидії інформаційним загрозам, проаналізовано національні та міжнародні підходи, механізми виявлення дезінформації в онлайн-середовищі та оцінено ефективність поточних рішень.

У ході дослідження було проаналізовано національний досвід України у сфері протидії інформаційним загрозам. Встановлено, що попри створення спеціалізованих державних структур та активну діяльність недержавних ініціатив, існують суттєві виклики, пов'язані з фрагментарністю зусиль, законодавчими обмеженнями (особливо щодо регулювання месенджерів) та недостатнім ресурсним забезпеченням. Це обмежує загальну ефективність протидії в умовах масштабної інформаційної агресії.

Розглянуто міжнародний досвід, зокрема підходи НАТО, які демонструють важливість комплексних стратегій, що охоплюють моніторинг, аналіз, превентивні комунікації, спростування та координацію зусиль між різними акторами. Вивчення цього досвіду дозволило ідентифікувати потенційно корисні моделі та практики для України.

Проаналізовано сучасні механізми виявлення фейкових повідомлень та дезінформації в онлайн-середовищі, включаючи традиційні методи (фактчекінг), автоматизовані (на основі ШІ/МН) та гібридні підходи. Розглянуто політики

модерації контенту провідних соціальних платформ (Meta, X/Twitter, YouTube, Telegram) та їхні обмеження.

Проведено емпіричну оцінку ефективності поточних рішень шляхом аналізу поширення деструктивного контенту на ключових онлайн-платформах. Результати дослідження підтвердили високу вразливість певних платформ, зокрема Telegram, до поширення неправдивої та маніпулятивної інформації, та вказали на недостатню ефективність існуючих механізмів контролю.

Таким чином, результати другого розділу підкреслюють необхідність розробки більш досконалих, комплексних та, зокрема, автоматизованих систем для ефективної протидії інформаційним загрозам, що створює логічний перехід до проектування запропонованого в даній роботі механізму.

РОЗДІЛ 3
РОЗРОБКА МОДЕЛІ АВТОМАТИЗОВАНОЇ СИСТЕМИ ПРОТИДІЇ ІНФОРМАЦІЙНІЙ ВІЙНІ

3.1. Обґрунтування необхідності впровадження автоматизованого механізму протидії дезінформації та визначення вимог до нього

З огляду на проаналізовані в попередніх розділах масштаби та специфіку інформаційних загроз, що постають перед Україною, нагальною потребою є розробка ефективних інструментів протидії. Цей розділ зосереджений на обґрунтуванні та проєктуванні моделі автоматизованого механізму, призначеного для виявлення та реагування на такі загрози.

В умовах перманентної інформаційної агресії, з якою стикається Україна, ефективна протидія дезінформації та іншим ворожим інформаційно-психологічним операціям набуває важливого значення. Існуючі зусилля різних державних структур, громадських організацій та медіа, хоч і є важливими, часто виявляються фрагментованими, недостатньо скоординованими та позбавленими системного підходу. Це призводить до розпорошення ресурсів, дублювання функцій, уповільненої реакції на нові виклики і, як наслідок, до зниження загальної ефективності протидії.

Для підвищення ефективності протидії, даний підрозділ обґрунтовує доцільність впровадження на базі існуючих державних структур, зокрема Центру протидії дезінформації, запропонованого автоматизованого механізму (системи). Цей механізм, має стати основним технологічним інструментом для розширеного моніторингу, інтелектуального аналізу інформаційних потоків, автоматизованої генерації попереджень та спростувань, а також для підтримки оперативного реагування на інформаційні загрози.

3.1.1 Необхідність посилення існуючих структур автоматизованим механізмом.

Складність сучасних інформаційних загроз та швидкість їх поширення обумовлюють необхідність переходу від переважно реактивних та частково ручних методів роботи до формування проактивної та технологічно підсиленої державної політики у сфері протидії дезінформації. Інтеграція запропонованого автоматизованого механізму в діяльність існуючих центрів, таких як ЦПД, може стати ключовим фактором такого переходу.

Переваги та функції вдосконаленої системи протидії на базі ЦПД з використанням запропонованого автоматизованого механізму полягають у можливості значно посилити спроможності держави. Такий механізм забезпечить централізацію та автоматизацію збору й обробки даних, дозволяючи аналітикам ЦПД працювати з більш повними та актуальними інформаційними потоками. Застосування ШІ дозволить проводити глибокий аналіз інформаційних кампаній, виявляти приховані зв'язки, ідентифікувати джерела та автоматично генерувати метадані. Це сприятиме прискоренню виявлення загроз та наданню оперативних даних для скоординованого реагування. Механізм також підтримуватиме проактивну роботу, допомагаючи виявляти ознаки підготовки ІПСО та автоматизуючи генерацію проектів попереджувальних повідомлень. Важливою функцією стане підтримка автоматизованої генерації проектів спростувань на основі виявлених фейків та зібраної доказової бази, що може бути використано для підвищення медіаграмотності населення та швидкого реагування.

Ключові завдання, що можуть бути ефективніше вирішені з допомогою запропонованого автоматизованого механізму в рамках ЦПД, охоплюють постійний та значно розширений моніторинг інформаційного простору. Механізм має сприяти глибокому автоматизованому аналізу виявлених загроз, включаючи ідентифікацію джерел, виявлення маніпулятивних технік, видобуток додаткових метаданих та оцінку впливу за допомогою ШІ. Орган,

використовуючи механізм, зможе швидше розробляти та реалізовувати заходи з нейтралізації загроз, спираючись на автоматично згенеровані проекти спростувань або попереджень, які потребуватимуть лише експертної верифікації. Це також посилить координацію діяльності та надання експертно-аналітичної підтримки.

3.1.2 Автоматизований механізм як невід'ємний інструмент державного органу

Ефективне виконання завдань ЦПД та інших уповноважених органів у сучасних умовах неможливе без впровадження передових технологічних рішень. Обсяги інформації в цифровому середовищі значні, а швидкість поширення дезінформації вимагає миттєвої реакції. Тому запропонований автоматизований механізм протидії інформаційним загрозам, що активно використовує технології штучного інтелекту для прискореного реагування та кращого виявлення загроз, стає не просто допоміжним, а критично важливим інструментом. Такий механізм, що інтегрує ШІ, машинне навчання (МН), обробку природної мови (NLP) та аналіз великих даних (Big Data), здатний суттєво посилити та оптимізувати спроможності існуючих державних структур.

Основні переваги автоматизації з акцентом на ШІ для державного органу типу ЦПД включають значне розширення охоплення моніторингу та прискорення виявлення загроз завдяки ідентифікації аномалій та ознак ІПСО в реальному часі за допомогою інтелектуальних алгоритмів. ШІ дозволяє здійснювати поглиблений аналіз для виявлення прихованих зв'язків між суб’єктами, джерелами та повідомленнями, автоматично видобувати більше релевантних метаданих та підвищувати об'єктивність оцінок при первинному аналізі контенту. Крім того, ШІ забезпечує автоматизацію рутинних завдань (збір, фільтрація, первинна класифікація даних), а також може автоматично генерувати проекти спростувань або попереджень на основі виявлених фейків та зібраної доказової бази, що дозволяє аналітикам зосередитися на верифікації та

прийнятті стратегічних рішень. Це сприяє створенню комплексної ситуаційної обізнаності та значно пришвидшує цикл реагування. Таким чином, розробка та впровадження такого складного автоматизованого механізму є важливою передумовою для підвищення ефективності протидії дезінформації.

Виходячи із специфіки сучасних загроз, до автоматизованого механізму висувається низка функціональних та нефункціональних вимог. Функціональні вимоги визначають ключові спроможності системи. Насамперед, система повинна забезпечувати автоматизований збір даних з широкого кола джерел, з гнучким налаштуванням параметрів та підтримкою різних типів контенту. Після збору дані мають проходити комплексну попередню обробку, включаючи очищення, нормалізацію (лематизацію/стемінг), автоматичне визначення мови, видобуток розширених метаданих за допомогою ШІ та транскрибування медіаконтенту.

Центральною функціональною можливістю системи має бути глибокий аналіз інформації для виявлення загроз з використанням ШІ. Це передбачає застосування передових алгоритмів машинного навчання для класифікації текстів (фейки, пропаганда, мова ворожнечі), аналізу тональності, тематичного моделювання та виявлення наративів. Критично важливою є спроможність системи до ідентифікації джерел та суб’єктів інформаційних впливів за допомогою ШІ (виявлення ботів, аналіз скоординованої поведінки, OSINT-інтеграція) та автоматичного виявлення ознак маніпуляцій у мультимедійному контенті (діпфейки, фотомонтаж). На основі комплексного аналізу ШІ має сприяти оцінці рівня загрози, пріоритезації інцидентів та, за можливості, автоматичній генерації проектів спростувань або попереджень на основі виявлених фейків та зібраної доказової бази, а також виділяти інші типи маніпуляцій для подальшого аналізу експертами.

Для забезпечення ефективної роботи операторів, система повинна мати розвинені інструменти звітності, візуалізації, управління інцидентами та сповіщення, що підтримують роботу з результатами аналізу ШІ та дозволяють оперативно реагувати. Необхідна система налаштовуваних сповіщень та

інструментарій для управління інформаційними інцидентами. Нарешті, система повинна включати централізоване сховище (архів та база знань) для всіх даних та результатів аналізу ШІ.

Нефункціональні вимоги визначають, якими характеристиками механізм повинен володіти. До них належать продуктивність та масштабованість, точність та надійність (особливо для моделей ШІ), безпека, зручність використання та адаптивність (включаючи можливість оперативного донавчання моделей ШІ).

Формулювання та подальше уточнення цих вимог є критично важливим етапом, що визначатиме успішність розробки та ефективність функціонування запропонованого автоматизованого механізму.

3.2. Теоретичні засади функціонування модулів автоматизованого механізму

Автоматизований механізм протидії інформаційним загрозам, що пропонується, є складною системою, яка функціонує на основі взаємодії низки спеціалізованих модулів (Додаток А). Кожен модуль виконує чітко визначену роль у загальному процесі обробки інформації – від її первинного збору до аналізу та надання інструментів для реагування[25]. Нижче наведено теоретичні засади функціонування кожного ключового модуля.

Модуль збору даних - первинна ланка системи, що відповідає за збір інформації з різноманітних зовнішніх джерел. Його основне призначення – забезпечення максимально повного та своєчасного надходження потенційно релевантних даних до системи. Модуль має бути спроектований для роботи з гетерогенними джерелами, включаючи структуровані (API, бази даних), напівструктуровані (веб-сторінки, RSS-канали) та неструктуровані дані (тексти, медіафайли). Фундаментальною задачею є не лише збір, але й первинна каталогізація джерел та забезпечення гнучкості конфігурації процесу збору.

Модуль попередньої обробки даних. Призначення цього модуля полягає у трансформації "сирих", часто зашумлених та неструктурованих даних,

отриманих від Модуля збору, у формат, придатний для подальшого аналізу та зберігання. Основи його роботи включають застосування методів очищення даних (видалення шумів, артефактів), нормалізації (приведення до стандартного вигляду), структурування (виділення значущих елементів) та додавання метаданих. Для текстових даних ключовими є операції лінгвістичної обробки, такі як токенізація, лематизація/стемінг, визначення мови. Для мультимедійних даних – транскрибування та базовий аналіз.

Центральне сховище даних - основа зберігання всієї інформації, що обробляється системою. Його призначення – безпечне, ефективне, та довготривалого зберігання даних різного типу та ступеня обробки, а також надання контрольованого доступу до них для всіх інших модулів. Сховище має підтримувати різні типи даних та забезпечувати механізми індексації, пошуку, управління життєвим циклом даних та їх резервного копіювання.

Аналітичне ядро - центральний інтелектуальний компонент системи, призначення якого – перетворення оброблених даних оперативну інформацію про інформаційні загрози. В основі його роботи лежать методи машинного навчання (ML), обробки природної мови (NLP), статистичного аналізу та інші техніки інтелектуального аналізу даних. Модуль має виконувати завдання класифікації контенту (виявлення фейків, пропаганди), аналізу тональності, ідентифікації та профілювання суб’єктів інформаційних впливів, виявлення аномалій та скоординованої поведінки, а також оцінки рівня загрози. Передбачається постійне навчання та адаптація аналітичних моделей.

Модуль управління інцидентами - Цей модуль забезпечить структурований процес обробки виявлених інформаційних загроз, перетворюючи їх на керовані інциденти. Його призначення – надати операторам інструменти для реєстрації, відстеження, аналізу, пріоритезації та координації дій щодо кожного інформаційного інциденту. Модуль має підтримувати життєвий цикл інциденту, дозволяти призначати відповідальних, документувати хід розслідування та вжиті заходи, а також накопичувати доказову базу.

Модуль сповіщень та попереджень - Призначення цього модуля – забезпечення своєчасного інформування визначених користувачів або систем про виявлення критичних інформаційних загроз, аномальної активності або інших значущих подій, що потребують негайної уваги. Теоретично, модуль має базуватися на системі правил та порогів, що налаштовуються, та інтегруватися з різними каналами комунікації для гарантованої доставки сповіщень відповідним адресатам.

Модуль візуалізації та взаємодії з користувачем - Цей модуль є точкою взаємодії між операторами системи та її автоматизованими компонентами. Його теоретичне призначення – надання інтуїтивно зрозумілого, функціонального та адаптивного графічного інтерфейсу для ефективного доступу до інформації, проведення аналізу, управління інцидентами, налаштування системи та отримання зворотної інформації. Важливими аспектами є представлення даних у наочній формі (дашборди, звіти, графіки, карти) та реалізація механізмів "людина-в-циклі" (HITL) для верифікації та донавчання аналітичних моделей[26].

Загальна модель інформаційного потоку та реагування. Інформація з зовнішніх джерел надходить до Модуля збору, проходить попередню обробку та зберігається у Центральному сховищі. Аналітичне ядро аналізує ці дані, виявляючи потенційні загрози. Виявлені загрози обробляються в Модулі управління інцидентами. На основі аналізу та класифікації загроз приймається рішення про тип реагування:

1. Превентивне реагування: Якщо виявлено ознаки підготовки ІПСО або зародження небезпечного наративу, Модуль управління інцидентами координує створення попереджувальних матеріалів. Модуль сповіщень (або спеціалізований інтерфейс в Модулі візуалізації) може бути використаний для їх поширення через офіційні канали комунікації з метою інформування громадськості до того, як шкідливий контент набуде значного поширення.

2. Активне реагування: Якщо виявлено вже поширений фейк або маніпуляцію, Аналітичне ядро надає доказову базу (ознаки діпфейку, аналіз

метаданих, невідповідності тощо). Модуль управління інцидентами координує підготовку та поширення спростувань через відповідні канали.

3.3. Архітектура та проектування ключових компонентів механізму автоматизованої протидії

Цей підрозділ детально розглядає, як кожен з теоретично обґрунтованих у підрозділі 3.2 модулів автоматизованого механізму виконуватиме свої функції на практиці. Для кожного модуля будуть описані ключові процеси, наведені приклади технологічних рішень та потенційні труднощі, що можуть виникнути на етапах розробки та експлуатації.

3.3.1. Модуль збору даних

Модуль збору даних реалізується як набір програмних компонентів (скриптів, сервісів), що взаємодіють із зовнішніми джерелами для отримання інформації.

Взаємодія з API: Для платформ, що надають API (наприклад, X API, YouTube Data API, Telegram Bot API для публічних каналів, API новинних агрегаторів), розробляються спеціалізовані клієнти. Ці клієнти автентифікуються (якщо потрібно), надсилають запити згідно з документацією API (наприклад, пошук за ключовими словами, отримання постів певних користувачів/каналів, збір коментарів) та обробляють отримані відповіді (зазвичай у форматі JSON або XML). Наприклад Python-скрипт, що використовує бібліотеку telethon або python-telegram-bot для підключення до Telegram API та збору повідомлень з визначеного списку публічних каналів кожні 10-15 хвилин.

Веб-скрейпінг: Для веб-сайтів (новинні портали, блоги, форуми), що не мають публічного API, застосовуються технології веб-скрейпінгу. Використовуються бібліотеки Python, такі як Requests (для виконання HTTP-

запитів) та BeautifulSoup або lxml (для парсингу HTML/XML-структури сторінок). Розробляються парсери, які за допомогою CSS-селекторів або XPath-виразів вилучають необхідний контент (заголовки, тексти статей, дати публікації, імена авторів, коментарі). Скрейпер на базі Scrapy (Python-фреймворк), що щогодини обходить головні сторінки визначену кількість новинних сайтів, переходить за посиланнями на нові статті та витягує їх повний текст і метадані.

Обробка RSS-каналів: Модуль періодично (кожні 5 хвилин) перевіряє оновлення у зазначених RSS-фідах новинних видань, завантажує та парсить XML-структуру для отримання посилань на нові матеріали, їх заголовків та коротких описів.

Динамічне виявлення джерел/публікацій: Для розширення охоплення можуть використовуватися алгоритми, які аналізують тренди в соціальних мережах (знову ж таки, через API, якщо платформа це дозволяє) або популярні запити в пошукових системах для виявлення нових потенційно значущих джерел або тем, що набирають популярності та потребують моніторингу.

Обробка скарг користувачів та сигналів від партнерів: Реалізується через захищену веб-форму на порталі системи, куди оператори, громадяни або довірені партнерські організації можуть надсилати URL-адреси, тексти або файли з підозрілим контентом для аналізу. Ці дані отримують спеціальний статус ("сигнал від користувача", "ручне введення", "пріоритетний сигнал від партнера").

Ручне введення даних операторами: Інтерфейс системи дозволяє операторам вручну додавати посилання на конкретні матеріали, завантажувати файли (наприклад, скріншоти, документи) або вводити іншу релевантну інформацію, отриману з неавтоматизованих джерел (наприклад, під час OSINT-розслідувань).

Зібрані дані, незалежно від джерела, уніфікуються (наскільки це можливо) за структурою (у форматі JSON-об'єктів з визначеним набором полів) та передаються до системи черг повідомлень (аналогічно до Apache Kafka або

RabbitMQ). Це забезпечує асинхронну обробку даних наступними модулями, відмовостійкість (якщо один з обробників тимчасово недоступний, дані залишаються в черзі) та можливість масштабування системи обробки. Кожному інформаційному об'єкту присвоюється унікальний ідентифікатор, фіксуються метадані про джерело, час збору, тип контенту тощо.

Потенційні ризики та шляхи їх мінімізації:

Одним із значних ризиків є динамічність веб-джерел, оскільки структура сайтів та умови використання API часто змінюються. Для мінімізації цього ризику пропонується впровадження системи моніторингу працездатності скрейперів та API-клієнтів, використання більш гнучких методів парсингу (наприклад, на основі візуального аналізу сторінки або машинного навчання для ідентифікації блоків контенту), а також регулярне оновлення та підтримка коду збирачів.

Іншою проблемою є заходи протидії скрейпінгу, такі як CAPTCHA або блокування за IP-адресою. Шляхами мінімізації цього можуть бути використання ротації IP-адрес через пули проксі-серверів, застосування сервісів розгадування CAPTCHA (з урахуванням вартості та етичних аспектів), емуляція поведінки реального користувача за допомогою headless browsers, а також пріоритезація джерел з офіційними API, де це можливо.

Обмеження API також становлять виклик, оскільки більшість платформ встановлюють ліміти на кількість запитів. Для подолання цього необхідно ретельно вивчати документацію API, дотримуватися встановлених лімітів, реалізовувати механізми поступового збільшення навантаження (exponential backoff) при отриманні помилок про перевищення лімітів та оптимізовувати запити для отримання максимальної кількості даних за один раз.

Забезпечення належної якості, повноти та достовірності даних є ще одним важливим аспектом, оскільки саме на основі цих даних прийматимуться подальші аналітичні рішення.

Для мінімізації ризиків, пов'язаних з низькоякісними або недостовірними даними, ключовим рішенням є розробка та інтеграція системи оцінки надійності

джерел інформації. Така система призначена для автоматизованої та напівавтоматизованої класифікації та ранжування джерел (веб-сайтів, акаунтів у соціальних мережах, медіа-ресурсів) за рівнем їхньої ймовірної достовірності та об'єктивності. Основними цілями такої системи є фільтрація інформаційного шуму, пріоритезація даних для аналізу, виявлення потенційно зловмисних суб’єктів та надання аналітикам контексту про джерело.

Оцінка може базуватися на аналізі широкого спектру характеристик: метаданих джерела (вік домену/акаунта, прозорість інформації про власника, наявність SSL-сертифіката, редакційна політика), характеристик контенту (історична точність попередніх публікацій, стиль викладу, наявність ознак сенсаційності чи мови ворожнечі, використання посилань на авторитетні першоджерела, оригінальність контенту), мережевої репутації (зворотні посилання, активність та характер взаємодії в соціальних мережах, згадки в "чорних" чи "білих" списках фактчекінгових організацій) та поведінкових індикаторів (аномальна частота публікацій, ознаки скоординованої поведінки). Методи оцінки можуть включати експертні системи на основі правил, моделі машинного навчання (навчені на розмічених даних про надійні та ненадійні джерела) або гібридні підходи. Важливо, щоб рейтинг надійності джерела динамічно оновлювався.

Результатом роботи такої системи є присвоєння кожному джерелу рейтингу або категорії надійності (наприклад, "високий рівень довіри", "сумнівне джерело", "дезінформаційне джерело") та формування детального профілю джерела. Ця інформація використовується для фільтрації вхідного потоку даних, встановлення пріоритетів при аналізі, а також слугує одним із факторів при оцінці загального рівня інформаційної загрози. Окрім системи оцінки джерел, важливим є впровадження механізмів валідації та верифікації самих зібраних даних, використання декількох джерел для перехресної перевірки інформації та чітке документування походження й метаданих кожного інформаційного об'єкта (рисунок 3.1).

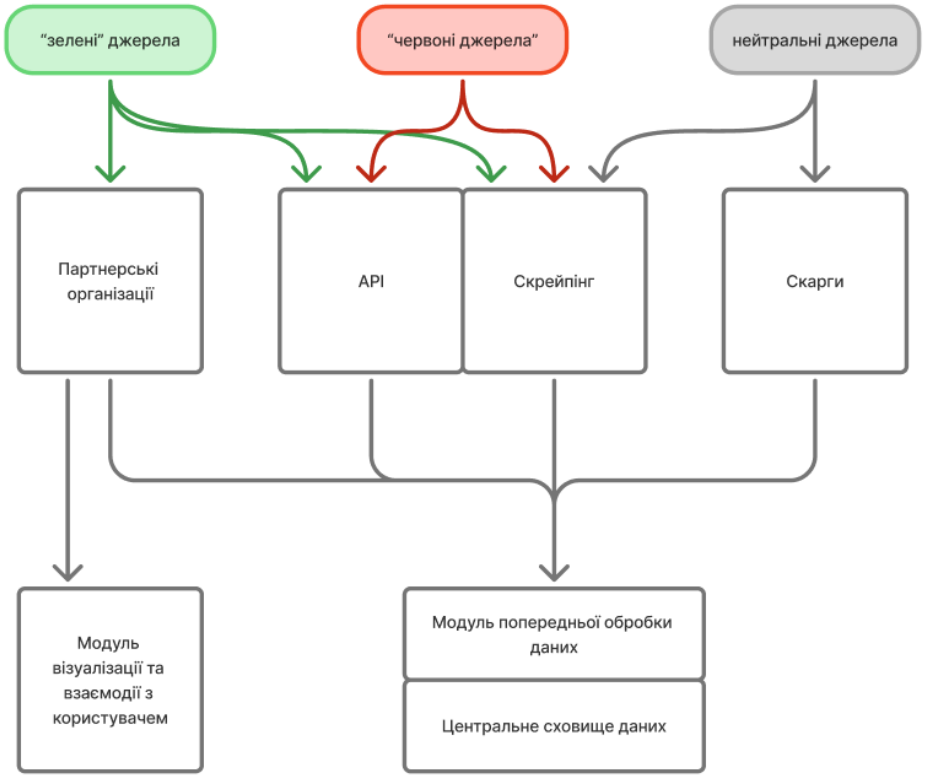


Рисунок 3.1 - Блок-схема роботи модуля збору даних

Значні обсяги даних (Big Data), що генеруються під час моніторингу, створюють навантаження на систему. Мінімізувати це можна шляхом використання масштабованих технологій зберігання та обробки даних (наприклад, розподілені файлові системи та хмарні сервіси) та впровадження політик управління життєвим циклом даних, включаючи архівування та видалення застарілої інформації .

Нарешті, необхідно враховувати правові та етичні аспекти збору даних. Для цього слід розробити та дотримуватися чіткої політики збору даних, що відповідає чинному законодавству (про захист персональних даних, авторське право) та умовам використання онлайн-платформ, а також застосовувати анонімізацію персональних даних, де це можливо та доцільно, і отримувати згоду на обробку даних, якщо це передбачено.

3.3.2. Модуль попередньої обробки даних

Цей модуль отримує "сирі" дані від Модуля збору та виконує їх підготовку для подальшого аналізу та зберігання, перетворюючи їх на структурований та очищений формат.

Очищення тексту: Для текстових даних використовуються регулярні вирази (re в Python) та спеціалізовані бібліотеки (наприклад, BeautifulSoup для видалення залишків HTML-тегів, якщо вони не були повністю видалені скрейпером) для видалення технічних артефактів, скриптів, стилів, спеціальних символів, надлишкових пробілів, емодзі (якщо вони не є предметом аналізу). Видалення всіх тегів <script>...</script> та <style>...</style> з HTML-фрагмента. Ідентифікація та видалення (або позначення) дублікатів повідомлень здійснюватися за допомогою хешування повного тексту (наприклад, MD5, SHA256).

Нормалізація тексту: Текст приводиться до єдиного регістру. Можуть застосовуватися алгоритми для виправлення поширених одруків (на основі словників). Уніфікується написання дат, чисел, абревіатур.

Визначення мови: Використовуються бібліотеки типу langdetect, fastText або спеціалізовані моделі для визначення мови кожного текстового фрагмента. Приклад: якщо система аналізує переважно український контент, повідомлення, визначене як англомовне, може бути оброблене окремим конвеєром.

Лематизація/Стемінг: Для української мови ефективно використовуються бібліотеки pymorphy2 або моделі з spaCy чи Stanza для приведення слів до їхньої базової словникової форми (леми)[27]. Це важливо для зменшення розмірності простору ознак та уніфікації слів перед аналізом (наприклад, "новина", "новини", "новиною" -> "новина").

Видалення стоп-слів - процедур попередньої обробки, що полягає у фільтрації високочастотних слів, які зазвичай не несуть семантичного навантаження (наприклад, прийменники, сполучники, частки, деякі займенники). Метою цієї операції є зменшення розмірності простору ознак,

підвищення обчислювальної ефективності наступних алгоритмів обробки природної мови (зокрема, класифікації текстів, тематичного моделювання, інформаційного пошуку). Процес зазвичай включає токенізацію вхідного тексту з подальшим порівнянням кожного токена з попередньо визначеним, мовоспецифічним списком стоп-слів (наприклад, наданим бібліотеками типу NLTK) та їх вилученням..

Видобуток метаданих та ознак: За допомогою моделей розпізнавання іменованих сутностей (NER) з spaCy або Stanza вилучаються особи, організації, локації, дати, грошові суми тощо. Можуть вилучатися ключові фрази (як це реалізовано в алгоритмі RAKE або на основі частотних характеристик TF-IDF). Ці видобуті сутності та фрази зберігаються як окремі атрибути інформаційного об'єкта.

Транскрибування та аналіз медіа: Аудіофайли та аудіодоріжки відеофайлів передаються до сервісів автоматичного розпізнавання мови (ASR) типу OpenAI Whisper, Google Cloud Speech-to-Text для перетворення на текст. Отриманий текст потім проходить аналогічну обробку, як і первинні текстові дані. Зображення проходять через базовий аналіз метаданих (EXIF), а також через моделі оптичного розпізнавання символів (OCR) для вилучення тексту з зображень.

Фільтрація непотрібної інформації: На основі попередньо визначених правил (наприклад, мінімальна/максимальна довжина тексту, наявність/відсутність певних ключових слів, відповідність попередньо визначеним тематикам, відсутність ознак спаму – велика кількість посилань, спецсимволів) більша частина інформації буде відфільтрована.

Результатом роботи модуля є потік оброблених, збагачених та структурованих даних (наприклад, у вигляді JSON-об'єктів з полями для очищеного тексту, лематизованих токенів, списку іменованих сутностей, визначеної мови, тональності тощо), готових для зберігання в аналітичних сховищах та для застосування складних аналітичних алгоритмів наступними модулями (рисунок 3.2).

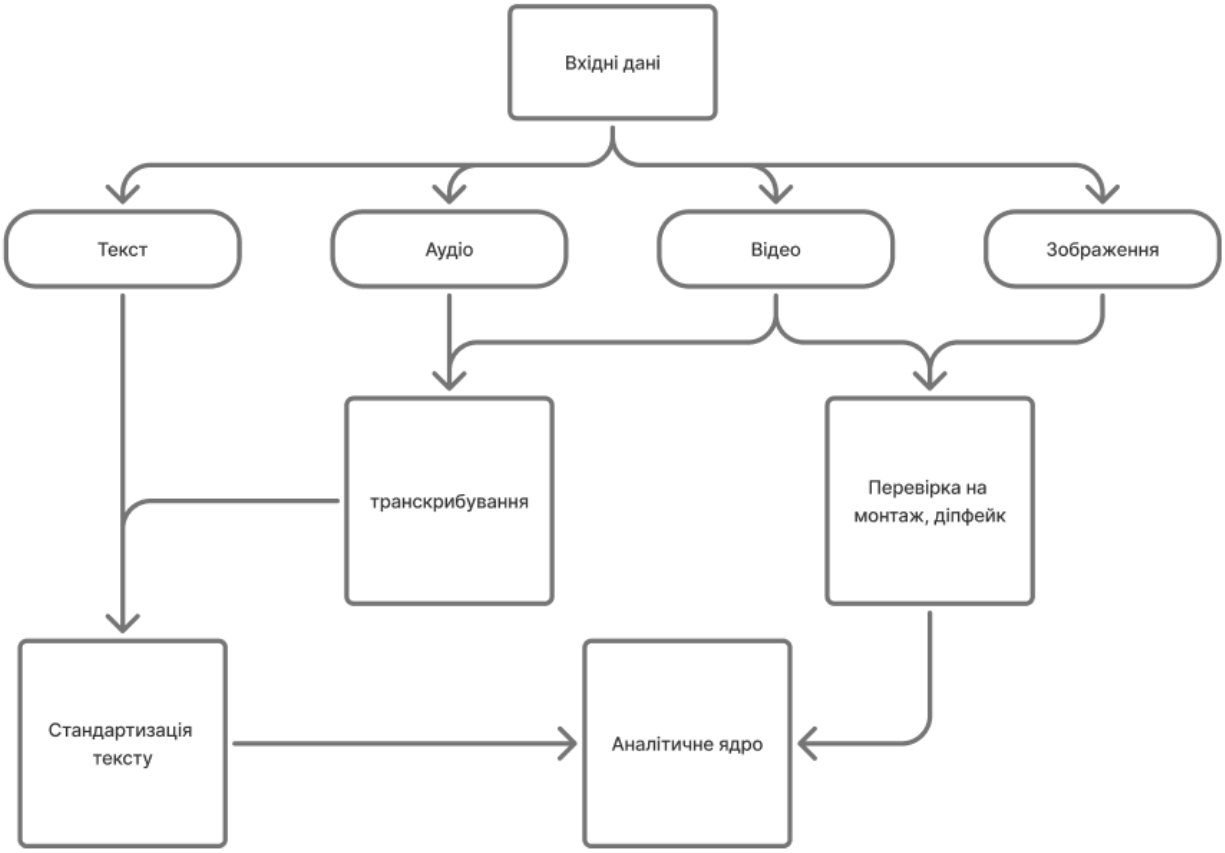


Рисунок 3.2 - Блок-схема роботи модуля попередньої обробки даних.

Низька якість лінгвістичних інструментів для української мови, порівняно з більш поширеними мовами, може стати перешкодою. Для мінімізації цього ризику слід розглянути можливість створення та донавчання власних моделей на специфічних для завдання даних, або ж комбінувати декілька інструментів для підвищення надійності результатів.

Неоднозначність природної мови, зокрема сарказм та іронія, важко піддається автоматичній обробці. На поточному етапі важливо усвідомлювати ці обмеження. Для більш глибокого аналізу варто покладатися на експертну оцінку у сумнівних випадках.

Існує ризик втрати важливого контексту при надто агресивному очищенні тексту. Щоб запобігти цьому, слід застосовувати збалансований підхід до очищення, зберігати "сирі" дані для можливості повернення до оригіналу.

Нарешті, адаптація до специфічного сленгу та неологізмів, що швидко з'являються в інформаційному просторі, є постійним викликом. Це вимагає регулярного оновлення словників та лінгвістичних моделей, а також використання методів, що здатні обробляти слова, відсутні у словнику (Out-Of-Vocabulary words), на основі символьних N-грам або можливостей сучасних трансформерних моделей.

3.3.3. Центральне сховище даних

Центральне сховище даних реалізується як комплекс взаємопов'язаних систем зберігання, кожна з яких оптимізована для певних типів даних та завдань доступу.

Data Lake: Сирі дані, отримані від Модуля збору (оригінальні тексти, HTML-файли, зображення, відео, аудіозаписи), зберігаються в їхньому первинному незмінному вигляді (аналогічно HDFS або AWS S3). Це дозволяє повторну обробку за новими алгоритмами. Нариклад: HTML-код веб-сторінки статті зберігається повністю для можливості перевірки джерела або вилучення додаткових метаданих у майбутньому. Дані тут можуть зберігатися у форматах Parquet або ORC для ефективної обробки великих обсягів.

Реляційні бази даних (RDBMS): Використовуються для зберігання структурованої інформації: метадані про зібрані об'єкти (URL, час збору, джерело, автор, кількість лайків), конфігурації системи, профілі користувачів, словники, класифікатори, результати агрегації..

База знань: Реалізується як інтегрована частина сховища, що може використовувати комбінацію вищезазначених технологій. Вона містить верифіковану інформацію про відомі фейки, спростування, профілі дезінформаційних суб’єктів, їхні типові тактики, техніки та процедури (TTPs), а також успішні контрзаходи. Запис у базі знань описує конкретну дезінформаційну кампанію, її цілі, методи та посилання на матеріали, що її спростовують. Вона також має містити шаблони повідомлень для спростувань та

попереджень, дані про ефективність різних каналів поширення контр-інформації та профілі цільових аудиторій.

Модуль забезпечує API для доступу до даних для інших компонентів системи, а також механізми управління правами доступу.

Значні обсяги, швидкість та різноманітність даних (Big Data) є ключовим викликом, що вимагає ретельного планування архітектури сховища. Для його подолання слід використовувати відповідні технології (аналогічні Data Lakes, NoSQL, розподілені системи), впроваджувати політики управління життєвим циклом даних (архівування, видалення неактуальної інформації) та оптимізовувати запити до даних.

Підтримка узгодженості даних (Data Consistency) між різними типами сховищ може бути складною. Тому пропонується використання механізмів синхронізації даних, застосуванням підходів типу "єдиного джерела правди" (Single Source of Truth) для ключових даних.

Складність проектування ефективної схеми даних та моделей знань, особливо для NoSQL баз даних та Бази Знань, вимагає залучення експертів з моделювання даних. Ітеративний підхід до проектування з можливістю еволюції схеми та використання стандартних онтологій, де це можливо, може спростити це завдання.

Сховище даних має бути готовим до еволюції вимог та технологій. Побудова гнучкої та розширюваної архітектури, використання відкритих стандартів та технологій з активною спільнотою, а також регулярний перегляд та оновлення технологічного стеку допоможуть підтримувати систему актуальною.

Нарешті, забезпечення високої продуктивності запитів до великих обсягів даних є критичним для операційної ефективності. Це досягається шляхом ефективного індексування даних, оптимізації SQL та NoSQL запитів, використання спеціалізованих аналітичних баз даних або пошукових рушіїв, а також кешування часто використовуваних даних.

3.3.4. Аналітичне ядро

Аналітичне ядро є ключовим компонентом, що перетворює оброблені дані на корисні знання та сигнали про загрози. Його робота базується на конвеєрі аналітичних задач.

Класифікація контенту: На вхід подаються оброблені тексти. Моделі машинного навчання (наприклад, навчений на попередньо розмічених даних SVM, логістична регресія, або більш складні трансформерні моделі типу BERT, RoBERTa, адаптовані для конкретного завдання класифікації) присвоюють тексту одну або кілька міток ("фейк", "пропаганда", "мова ворожнечі", "нейтрально", "клікбейт") разом із показником впевненості (confidence score). Приклад: стаття про політичну подію може бути класифікована як "пропаганда" з впевненістю 0.85, а коментар до неї – як "мова ворожнечі" з впевненістю 0.92.

Аналіз тональності (сентимент-аналіз): Тексти аналізуються для визначення їх емоційного забарвлення. Це може бути реалізовано за допомогою лексикон-орієнтованих підходів (використання словників слів з попередньо присвоєними балами тональності, аналогічно до SentiWordNet, VADER) або моделей МН (наприклад, рекурентні нейронні мережі LSTM/GRU або трансформери, навчені на текстах з відомою тональністю). Результатом є оцінка тональності (“позитивна”, “негативна”, “нейтральна”) та, можливо, інтенсивності емоцій.

Тематичне моделювання (Topic Modeling): Застосовуються алгоритми (наприклад, LDA (Latent Dirichlet Allocation), NMF (Non-negative Matrix Factorization) або більш сучасні підходи на основі нейронних мереж[28,29], такі як Top2Vec або BERTopic) до великих корпусів текстів для автоматичного виявлення прихованих тем та наративів. Кожна тема представляється набором ключових слів. Приклад: аналіз тисяч новинних статей за тиждень може виявити домінуючі теми, такі як "економічна ситуація в країні", "міжнародні переговори щодо безпеки", "нові соціальні ініціативи уряду", "обговорення законопроекту X".

Ідентифікація та профілювання джерел та суб’єктів інформаційних впливів. Виявлення ботів/тролів та скоординованої поведінки (CIB): Аналізуються поведінкові патерни акаунтів (аномально висока частота публікацій, нетиповий час активності, різка зміна тематики, велика кількість однакових або схожих повідомлень, аномальна кількість підписників/підписок, відсутність особистої інформації)[30], характеристики контенту (використання шаблонних фраз, надмірна кількість посилань на сумнівні ресурси, граматичні/лексичні помилки, характерні для машинних перекладів) та мережеві зв'язки (участь у скоординованих кампаніях лайків/репостів, швидке поширення інформації через мережу пов'язаних акаунтів). Приклад: група з 50 новостворених акаунтів у Twitter, які одночасно починають публікувати однакові негативні коментарі під постами певного політика, може бути позначена як підозріла на бот-мережу. Аналіз соціальних мереж (SNA): Будуються графи взаємодій (хто кого цитує, репостить, коментує, згадує). Застосовуються алгоритми для виявлення спільнот (community detection), центральних вузлів (впливових суб’єктів, інфлюенсерів), мостів між спільнотами, що можуть відігравати роль у поширенні інформації між різними групами. Приклад: виявлення невеликої групи тісно пов'язаних між собою блогерів, які синхронно поширюють та підсилюють певний дезінформаційний наратив. OSINT-техніки: Використання відкритих джерел та спеціалізованих інструментів для збору додаткової інформації про підозрілі акаунти[31], веб-сайти, організації для спроби їх атрибуції.

Виявлення ознак підготовки ІПСО: Модуль аналізує слабкі сигнали, нетипову активність, координацію між маловідомими джерелами, тестування певних наративів на обмеженій аудиторії, що може свідчити про підготовку масштабної кампанії. Приклад: різке зростання кількості новостворених акаунтів у соцмережі, що починають поширювати схожі повідомлення на певну тему, може бути сигналом для превентивного аналізу.

Аналіз вразливостей цільових аудиторій: На основі аналізу публічних даних та соціально-демографічних характеристик, ядро може допомагати

визначати, які групи населення є найбільш вразливими до певних типів дезінформації, що важливо для таргетування превентивних заходів.

Формування доказової бази для спростувань: Автоматичне виявлення та документування ознак маніпуляції (наприклад, невідповідність метаданих зображення його змісту, використання старих фотографій у новому контексті, ознаки редагування відео, лінгвістичні маркери пропаганди). Приклад: система виявляє, що фотографія, яка використовується у новині про нещодавню подію, насправді була зроблена кілька років тому в іншому місці.

Виявлення ознак маніпуляцій з медіаконтентом: Зображення та відео аналізуються спеціалізованими CNN-моделями, навченими розпізнавати артефакти, що залишаються при створенні діпфейків, або невідповідності в освітленні, тінях, текстурах.

Оцінка рівня загрози та пріоритезація: Розробляється система скорингу, яка на основі комбінації факторів (тип контенту, тональність, масштаб поширення, авторитетність джерела, ознаки координації, потенційна вразливість цільової аудиторії) присвоює кожному інформаційному об'єкту або інциденту числовий показник загрози.

Моделі МН постійно донавчаються (MLOps) на нових даних та за результатами верифікації операторами (HITL), що забезпечує їх адаптацію до мінливого інформаційного середовища.

Блок схема роботи аналітичного ядра зображена на рисунку 3.3.

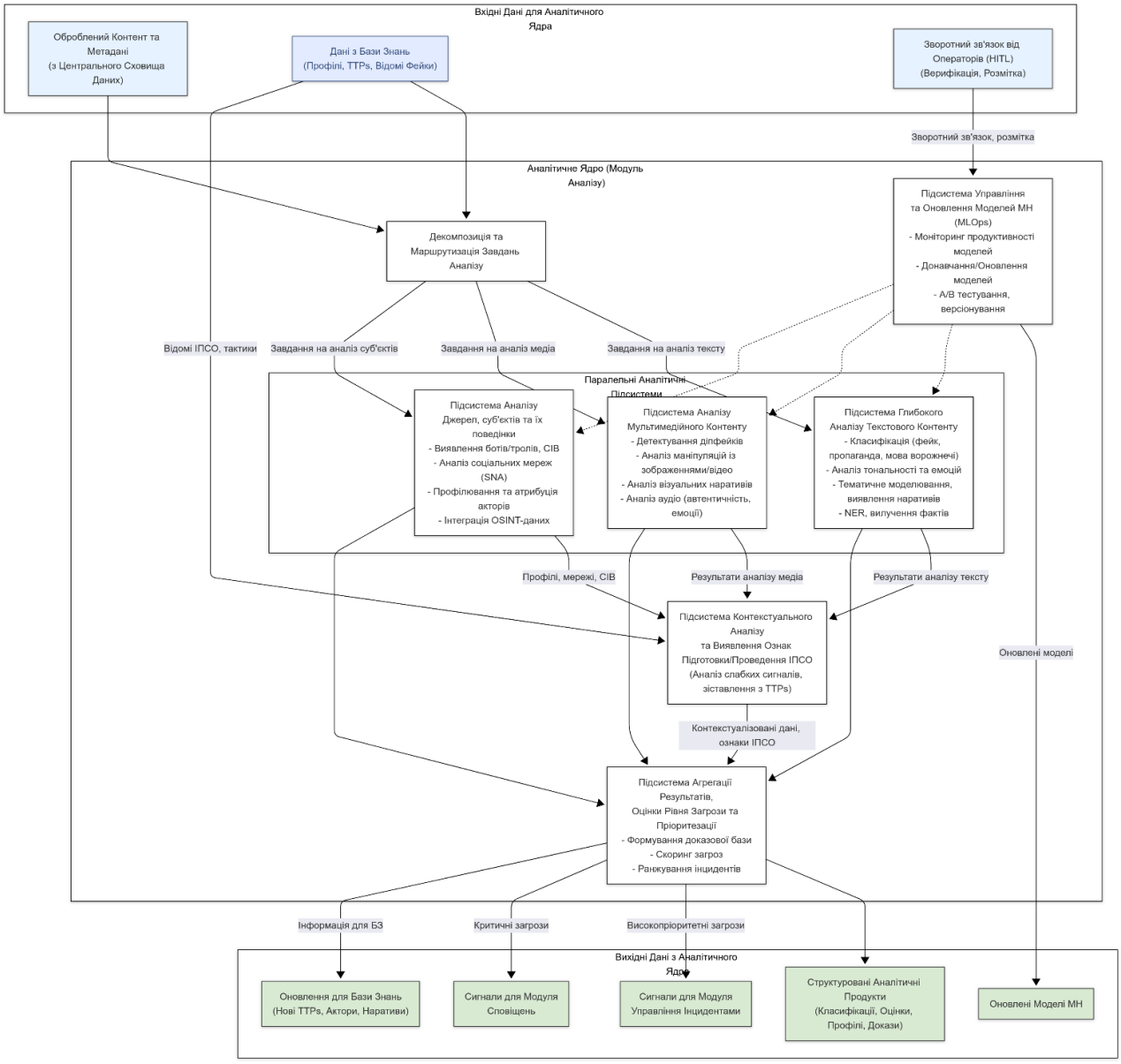


Рисунок 3.3 - Блок-схема роботи аналітичного ядра

Недостатня якість та кількість навчальних даних для моделей машинного навчання, особливо для специфічних типів загроз, є суттєвим ризиком. Необхідно інвестувати у створення та постійне оновлення якісних, збалансованих та репрезентативних розмічених наборів даних, використовувати техніки аугментації даних для штучного збільшення їх обсягу, а також застосовувати методи трансферного навчання (використання перед-навчених моделей) та навчання на слабких сигналах (weak supervision).

Постійна адаптація моделей до нових тактик дезінформації ("гонка озброєнь") є ще одним викликом, оскільки зловмисники безперервно

вдосконалюють свої методи. Щоб протистояти цьому, слід впроваджувати практики швидкого перенавчання та розгортання оновлених моделей, здійснювати постійний моніторинг ефективності моделей в реальних умовах та проводити дослідження нових типів загроз для розробки відповідних детекторів.

Високі вимоги до обчислювальних ресурсів для навчання та використання сучасних моделей МН можуть бути значним обмеженням. Шляхами оптимізації є застосування технік оптимізації моделей, використання хмарних обчислювальних ресурсів з можливістю гнучкого масштабування та розробка ефективних алгоритмів обробки даних.

Необхідність врахування контекстуальної залежності та культурних особливостей інформації є складним завданням для автоматичних систем. Для цього слід активно залучати експертів для інтерпретації результатів у складних, неоднозначних випадках.

Важливо підтримувати баланс між автоматизацією та експертною оцінкою. Автоматизовані системи мають розглядатися як інструменти підтримки прийняття рішень, а не як повна заміна експерта. Розробка ефективних інтерфейсів "людина-в-циклі" (HITL) для верифікації та корекції результатів є ключовою.

Визначення правильного порогу для превентивного реагування на основі ранніх ознак підготовки ІПСО є комплексним завданням, оскільки передчасне реагування може бути контрпродуктивним. Для цього пропонується розробка багаторівневої системи оцінки достовірності таких сигналів та використання декількох незалежних індикаторів з обов'язковим залучення експертів для прийняття остаточного рішення про публічне попередження, при будь-якій не однозначності.

Нарешті, складність точного прогнозування розвитку та впливу ІПСО вимагає використання сценарного моделювання, аналізу історичних даних про схожі кампанії та розробки моделей, що враховують динаміку поширення інформації та характеристики цільових аудиторій.

3.3.5. Модуль управління інцидентами

Цей модуль є центральним вузлом для організації роботи аналітиків з виявленими інформаційними загрозами, забезпечуючи структурований підхід до їх обробки та реагування.

Створення та реєстрація інциденту: Інциденти можуть створюватися автоматично (на основі сигналів від Аналітичного ядра), напівавтоматично (оператор підтверджує створення) або вручну. Кожному інциденту присвоюється унікальний ID, фіксується час, джерело сигналу.

Збагачення інциденту даними: Автоматично підтягується вся релевантна інформація з Центрального сховища та Аналітичного ядра. Оператори можуть додавати нотатки, файли, встановлювати зв'язки з іншими інцидентами/об'єктами Бази знань.

Призначення, пріоритезація та відстеження статусу: Інциденти призначаються аналітикам/групам, встановлюється пріоритет та терміни. Статус інциденту відстежується ("Новий", "В аналізі", "Потребує реагування", "Реагування триває", "Закритий – спростовано/заходи вжито", "Хибне спрацювання").

Планування та координація заходів реагування: Модуль надає інструменти для планування інформаційних кампаній: визначення цілей, аудиторій, ключових повідомлень, каналів поширення, відповідальних, термінів. Приклад превентивного реагування: на основі сигналу про підготовку ІПСО щодо виборів, створюється план превентивної кампанії з роз'ясненням маніпуляцій. Приклад активного реагування: для спростування фейку про мобілізацію створюється завдання з підготовки прес-релізу, постів для соцмереж, відео-коментаря.

Взаємодія з підрозділами комунікацій: Передача підготовлених та затверджених матеріалів (попереджень, спростувань) для публічного поширення.

Моніторинг та оцінка ефективності реагування: Інтеграція з іншими модулями для відстеження реакції аудиторії на контр-інформацію, аналізу зміни тональності, зменшення поширення фейку.

Ведення історії та оновлення Бази знань: Усі дії протоколюються. Після закриття інциденту ключова інформація (опис загрози, методи, результати, ефективність контрзаходів) структурується та додається до Бази знань.

Значна кількість автоматично генерованих інцидентів може призвести до перевантаження аналітиків, особливо якщо Аналітичне ядро має високий рівень хибних спрацювань. Необхідна ефективна система автоматичної пріоритезації інцидентів, чіткі критерії для їх створення, і в першу чергу можливість групової обробки схожих інцидентів або їх автоматичного закриття за певними умовами.

Складність координації між різними підрозділами або відомствами під час реагування на масштабні інформаційні атаки є ще одним викликом. Вбудовані в модуль інструменти для спільної роботи, призначення завдань та відстеження їх виконання, а також чітко визначені протоколи міжвідомчої взаємодії допоможуть подолати цю проблему.

Забезпечення оперативності узгодження та публікації контр-матеріалів є критичним, оскільки бюрократичні процедури можуть нівелювати швидкість реагування. Для цього слід розробити прискорені процедури узгодження для критичних випадків та попередньо затвердити шаблони й ключові повідомлення для типових сценаріїв.

Складність об'єктивного вимірювання ефективності заходів реагування вимагає визначення чітких Ключових Показників Ефективності (KPI) для кампаній реагування та використання комбінації кількісних (охоплення аудиторії, залученість, зміна частоти згадок фейку/спростування) та якісних (аналіз тональності публічних обговорень, проведення опитувань громадської думки) методів оцінки.

Існує також ризик "ефекту зворотної дії" (backfire effect), коли невдало сформульоване спростування може посилити віру в дезінформацію. Мінімізувати цей ризик можна шляхом ретельної підготовки спростувань з

урахуванням психологічних аспектів сприйняття інформації, фокусу на наданні перевірених фактів та альтернативних наративів, а не лише на прямому запереченні, а також можливого тестування ключових повідомлень на невеликих фокус-групах перед їх масовим поширенням.

3.3.6. Модуль сповіщень та попереджень

Цей модуль слугує системою раннього попередження та оперативного інформування як внутрішніх користувачів системи, так і, зовнішніх аудиторій.

Джерела тригерів для сповіщень: Сповіщення генеруються на основі сигналів від Аналітичного ядра (перевищення порогу загрози, виявлення скоординованої атаки), змін статусів у Модулі управління інцидентами (призначення інциденту, наближення дедлайну) або ручних тригерів від операторів.

Налаштування правил та шаблонів: Система дозволяє створювати та гнучко налаштовувати правила для генерації сповіщень (на основі типу загрози, джерела, ключових слів, швидкості поширення). Доступні шаблони для різних типів сповіщень. Приклад правила: "Якщо виявлено >100 повідомлень з ключовим словом 'прорив' у регіоні N за останню годину з недовірених джерел + класифікація 'потенційний фейк' >0.8 впевненості => термінове сповіщення групі X".

Визначення отримувачів та каналів доставки: Для кожного типу сповіщення визначаються групи отримувачів (аналітики, керівники, чергові служби, зовнішні партнери) та канали доставки (email, SMS, push-сповіщення, захищені месенджери, системи оперативного оповіщення).

Для критичних сповіщень можуть бути налаштовані рівні ескалації: якщо немає реакції протягом визначеного часу, сповіщення пересилається на вищий рівень.

Формування та підтримка поширення публічних попереджень та спростувань: На основі рішення з Модуля управління інцидентами, модуль може

сприяти поширенню попереджень або спростувань для громадськості через інтеграцію з офіційними сайтами, сторінками в соцмережах, розсилкою для ЗМІ. Приклад: генерація драфту повідомлення для Facebook-сторінки держоргану про фішингову кампанію з рекомендаціями.

Журналювання та відстеження: Усі відправлені сповіщення, їх отримувачі, час, канали та статуси доставки детально логуються.

Значна кількість некритичних або хибних сповіщень може спричинити "шум" та "втому від тривог" в операторів. Для запобігання цьому необхідне точне налаштування правил генерації сповіщень, їх ефективна пріоритезація, а також надання користувачам можливості (в межах політики безпеки) налаштовувати власні налаштування щодо отримання сповіщень.

Своєчасність та гарантованість доставки сповіщень є критичними, проте технічні проблеми з каналами зв'язку можуть спричинити затримки. Важливим є використання надійних каналів доставки, постійного моніторингу стану системи сповіщень та наявності резервних каналів зв'язку для критичних повідомлень.

Управління групами отримувачів та їхніми актуальними контактними даними може бути складним у великих організаціях. Інтеграція з корпоративними каталогами користувачів (LDAP, Active Directory), регулярна актуалізація контактної інформації та чіткі процедури управління групами розсилки допоможуть вирішити цю проблему.

Безпека каналів сповіщень є важливою при передачі чутливої інформації. Необхідно використовувати шифровані канали зв'язку та обмежувати обсяг чутливої інформації в самому тексті сповіщення, надаючи перевагу посиланням на захищені ресурси системи.

Вибір правильних каналів та формулювання повідомлень для публічного інформування вимагає залучення спеціалістів з комунікацій до розробки шаблонів та стратегій поширення, а також аналізу ефективності різних каналів для різних цільових аудиторій.

3.3.7. Модуль візуалізації та взаємодії з користувачем

Цей модуль є основним "робочим столом" для аналітиків, операторів та керівників, забезпечуючи їм доступ до всіх даних та функцій системи.

Дашборди (Інформаційні панелі): Надають узагальнене візуальне представлення ключових показників та поточної ситуації в інформаційному просторі в режимі реального часу. Дашборди можуть бути кастомізовані під різні ролі користувачів. Приклад: дашборд для керівника може показувати загальну кількість виявлених загроз, їх розподіл за типами та регіонами; дашборд для аналітика – нові інциденти, активність джерел.

Система звітності: Дозволяє генерувати стандартизовані та спеціалізовані звіти в різних форматах (PDF, Excel), що включають текстовий опис та візуалізації.

Інструменти пошуку та фільтрації: Надають можливість глибокого пошуку інформації в Центральному сховищі за різними критеріями (ключові слова, дати, джерела, автори, теги, статус інциденту).

Інструменти візуалізації даних: Інтерактивні інструменти для візуального представлення даних: географічні карти поширення впливів, графіки динаміки тем/тональності, мережеві діаграми зв'язків між акторами/джерелами/контентом. Приклад: граф, що показує, як акаунти в соцмережі синхронно поширюють посилання на фейковий сайт.

Інтерфейс управління інцидентами: Дозволяє операторам переглядати деталі інцидентів, змінювати статус, призначати відповідальних, додавати коментарі, завантажувати докази.

Інтерфейс "людина-в-циклі" (HITL): Спеціалізований інтерфейс для експертів, який дозволяє верифікувати результати роботи моделей МН, розмічати дані для навчання/донавчання моделей, надавати зворотний зв'язок.

Інструменти для підготовки матеріалів реагування: Шаблони для спростувань/попереджень, можливість додавання візуальних доказів (порівняння оригіналу та діпфейку).

Візуалізація ефективності контрзаходів: Дашборди, що показують динаміку поширення дезінформації до та після спростування, охоплення аудиторії контр-наративами.

Інтерфейс для координації публічних комунікацій: Відстеження статусу публікації попереджень/спростувань через офіційні канали.

Управління користувачами та правами доступу: Адміністративний інтерфейс для створення/управління обліковими записами, призначення ролей та налаштування прав доступу.

Для ефективної візуалізації великих та складних наборів даних слід використовувати сучасні бібліотеки та інструменти для інтерактивної візуалізації, розробляти спеціалізовані візуалізації для конкретних аналітичних завдань (наприклад, теплові карти для геоаналізу, часові шкали для динаміки) та надавати користувачам можливість налаштовувати параметри візуалізації.

Необхідно забезпечити гнучку кастомізацію інтерфейсу під потреби різних користувачів. Це може бути досягнуто шляхом розробки системи віджетів, що налаштовуються, можливості збереження користувацьких налаштувань дашбордів та звітів, а також впровадження рольової моделі доступу до різного функціоналу.

Безпека веб-інтерфейсу та захист від типових веб-вразливостей є критично важливими. Для цього слід дотримуватися стандартів безпечної розробки (OWASP Top 10), проводити регулярне тестування на проникнення та використовувати сучасні веб-фреймворки з вбудованими механізмами захисту.

Нарешті, навчання користувачів ефективному використанню всіх можливостей системи є важливим для її успішного впровадження. Розробка детальної та доступної документації користувача, створення інтерактивних навчальних посібників та відеоуроків, а також проведення регулярних тренінгів для операторів та аналітиків допоможуть подолати цей виклик.

3.4. Практична реалізація та тестування компонента механізму

Після теоретичного обґрунтування концепції автоматизованого механізму протидії інформаційним загрозам, визначення вимог до нього, а також детального проектування його архітектури та ключових компонентів (висвітлених у підрозділах 3.1-3.3), даний підрозділ присвячений практичній реалізації запропонованих ідей. Це здійснюється через розробку, реалізацію та тестування програмного прототипу, що втілює ключові функції розробленого механізму, з акцентом на автоматизації процесу сповіщення про виявлені загрози.

Практична реалізація дозволяє не лише продемонструвати життєздатність та функціональність теоретичних напрацювань, але й виявити потенційні технологічні та методологічні виклики на ранніх етапах, оцінити ефективність обраних підходів та отримати цінний емпіричний досвід для подальшого вдосконалення та підготовки до можливого повномасштабного розгортання системи.

Для практичної реалізації в рамках демонстрації було прийнято рішення зосередитись на створенні програмного прототипу, який інтегрує ключові функції Модуля збору даних, Модуля попередньої обробки даних, елементи Аналітичного ядра та Модуля сповіщень та попереджень .

Фундаментальна важливість обраних модулів: Збір даних, їх первинна обробка та базовий аналіз є обов'язковими етапами. Додавання автоматизованого сповіщення демонструє практичну спрямованість механізму на оперативне реагування.

Можливість демонстрації ключових технологій: Реалізація дозволяє застосувати веб-скрейпінг, взаємодію з API , базову обробку природної мови та розробку системи сповіщень.

Наочність та інтерпретованість результатів: Робота компонента зі збору, аналізу та автоматичного сповіщення про потенційні загрози дозволяє наочно продемонструвати його працездатність.

Основа для подальшого масштабування та розвитку: Розроблений прототип може слугувати основою для нарощування функціоналу, зокрема, вдосконалення аналітичних можливостей та розширення каналів сповіщення.

Обраний для реалізації прототип зосереджений на автоматизованому зборі текстового контенту з веб-сайтів та YouTube, їх попередній обробці, базовій класифікації та, у випадку виявлення потенційно маніпулятивного контенту, автоматичному надсиланні сповіщення через Telegram-бот.

Основні етапи розробки включали:

1) Визначення джерел даних та формату вхідних/вихідних даних: Джерела: список URL-адрес веб-сайтів (urls.txt) та YouTube (пошук за ключовими словами з keywords.txt). Вихідні дані: CSV-файл з результатами аналізу та сповіщення в Telegram.

2) Розробка функціоналу збору даних з веб-сайтів та YouTube: Створення функцій для отримання текстового контенту (заголовки, тексти статей, описи відео, коментарі).

3) Розробка модуля попередньої обробки тексту: Імплементація функції для очищення тексту та приведення його до нижнього регістру.

4) Розробка базового класифікатора контенту: Створення алгоритму для виявлення потенційно маніпулятивного контенту на основі аналізу присутності ключових слів/фраз з файлу keywords.txt.

5) Розробка Telegram-бота для сповіщень (notification_bot.py): Створення окремого скрипта для Telegram-бота, що приймає HTTP-запити та надсилає повідомлення у визначений чат. Використання Flask для створення простого веб-сервера.

6) Інтеграція аналізатора з Telegram-ботом: Модифікація основного скрипта ACIW_preview.py для надсилання HTTP POST-запитів до запущеного Telegram-бота у випадку виявлення потенційної загрози.

7) Створення основного керуючого скрипту та інтерфейсу для виведення результатів: Розробка головного скрипту ACIW_preview.py, що координує

роботу всіх функцій, та реалізація виведення результатів в консоль та збереження у CSV-файл.

Основним інструментарієм, мовами програмування та бібліотеками, що використовувалися для розробки прототипу, були:

- Мова програмування: Python версії 3.9.
- Бібліотеки для веб-скрейпінгу (збір даних з URL): Requests.
- Бібліотека для роботи з YouTube Data API v3: google-api-python-client.
- Бібліотеки для Telegram-бота та веб-сервера: python-telegram-bot, Flask.
- Бібліотеки для обробки тексту та даних: re (вбудована), Pandas.
- Середовище розробки (IDE): Visual Studio Code.

Серед ключових аспектів реалізації варто відзначити створення двох основних скриптів: ACIW_preview.py (аналізатор) та notification_bot.py (бот для сповіщень). Конфігурація джерел, ключових слів та токенів/ключів API винесена у відповідні константи або зовнішні файли.

Тестування розробленого програмного прототипу мало на меті комплексну перевірку його загальної працездатності, коректності виконання основних функцій збору та обробки даних, оцінку ефективності базового класифікатора, а також надійності системи автоматизованого сповіщення. Методика тестування передбачала кілька послідовних кроків. Спочатку проводилося тестування модулів збору даних шляхом перевірки коректності вилучення текстового контенту з визначених веб-сайтів та з YouTube, а також оцінювалася обробка можливих помилок API. Далі, модуль попередньої обробки даних тестувався на якість очищення тексту шляхом візуальної перевірки результатів. Після цього, ефективність базового класифікатора оцінювалася на невеликому, попередньо вручну розміченому наборі текстових фрагментів (близько 50 одиниць), для якого розраховувалися базові метрики точності та повноти. На завершення, перевірялася працездатність системи сповіщень через Telegram-бот: фіксувалося успішне надсилання повідомлень ботом у випадку виявлення програмним скриптом ACIW_preview.py контенту, класифікованого як "потенційно маніпулятивний".

Результати тестування показали, що модулі збору та попередньої обробки даних продемонстрували успішну та коректну роботу з тестовими джерелами інформації. Базовий класифікатор зміг ідентифікувати частину маніпулятивних фрагментів (рисунки 3.4 , 3.5), однак його ефективність значною мірою залежить від якості та повноти списку ключових слів, що використовуються для аналізу. Система сповіщень через інтегрованого Telegram-бота успішно виконувала свою функцію, надсилаючи повідомлення про виявлені потенційні загрози до вказаного чату (рисунок 3.6).

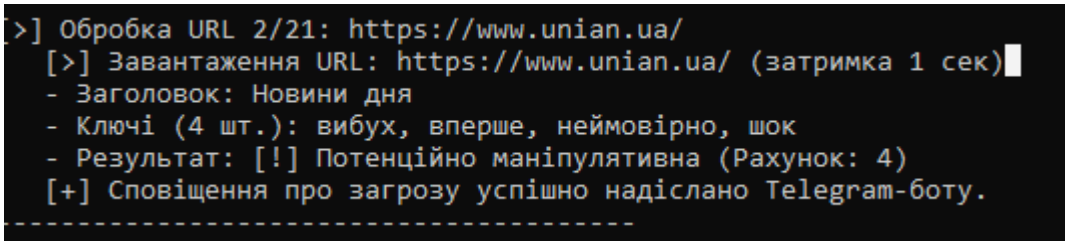


Рисунок 3.4 - Приклад консольного виводу ACIW_preview.py при аналізі URL адреси новинного сайту та спрацюванні класифікатора

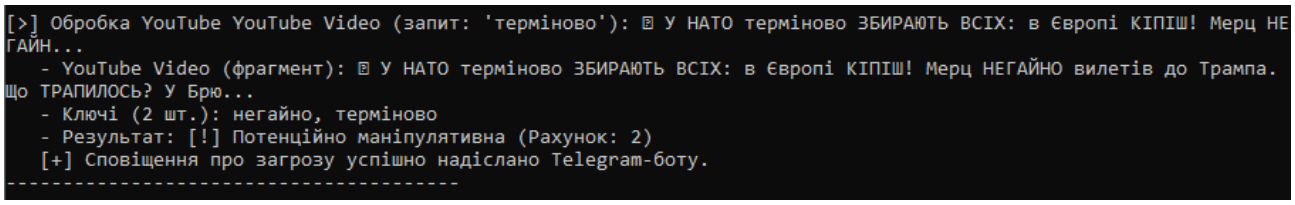


Рисунок 3.5 - Приклад консольного виводу ACIW_preview.py при аналізі даних з YouTube та спрацюванні класифікатора

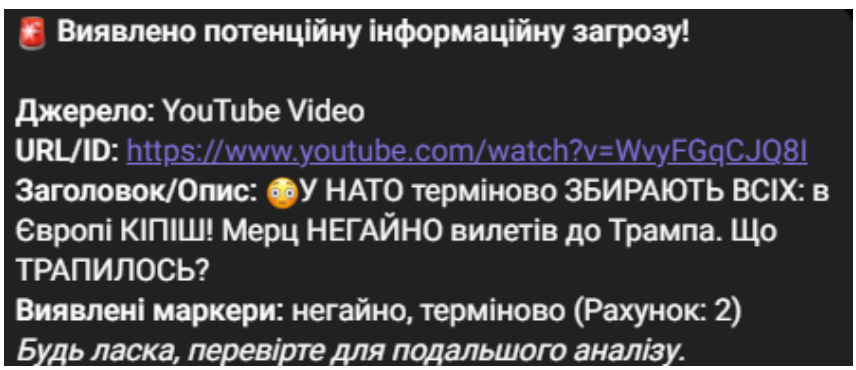


Рисунок 3.6 - Приклад сповіщення, отриманого від Telegram-бота notification_bot.py у чаті, про виявлену потенційну загрозу

Аналіз результатів та висновки з тестування свідчать, що розроблений програмний прототип успішно виконує покладені на нього базові функції зі збору даних, їх попередньої обробки, елементарної класифікації та автоматизованого сповіщення. Це підтверджує життєздатність обраного підходу на концептуальному рівні. Використання мови програмування Python та обраного набору бібліотек виявилося ефективним для реалізації поставлених завдань. Простий класифікатор на основі ключових слів має очікувано обмежену ефективність і потребує вдосконалення шляхом застосування методів машинного навчання. Інтеграція Telegram-бота для автоматизованих сповіщень є дієвим способом підвищення оперативності інформування про потенційні загрози. Практична реалізація підтвердила важливість якісної попередньої обробки текстових даних.

Отримані результати, а також виявлені обмеження простого підходу до класифікації контенту, чітко вказують на перспективні напрямки для подальшого суттєвого вдосконалення системи. Це, зокрема, стосується необхідності інтеграції більш досконалих та адаптивних алгоритмів машинного навчання для аналізу тексту, розширення та постійного оновлення бази знань про інформаційні загрози та тактики їх поширення, а також покращення механізмів адаптації системи до нових викликів в інформаційному просторі. Розроблений прототип є важливим практичним кроком на шляху до створення повноцінного та дієвого інструменту для підтримки діяльності спеціалізованого державного органу у сфері забезпечення інформаційної безпеки.

3.5. Перспективи розвитку механізму та його інтеграції в діяльність єдиного державного органу

Розробка та успішне тестування програмного прототипу ключових компонентів автоматизованого механізму протидії інформаційним загрозам, описані в попередньому підрозділі, є важливим початковим етапом. Вони

демонструють принципову життєздатність запропонованих рішень та закладають основу для подальшої роботи.

Однак, для створення повноцінного, високоефективного та стійкого до сучасних викликів інструменту, здатного стати надійною технологічною опорою для спеціалізованого державного органу, такого як Центр протидії дезінформації (ЦПД), необхідний подальший системний розвиток як самого механізму, так і розробка комплексної стратегії його глибокої інтеграції в операційні процеси та організаційну структуру відповідної державної інституції.

Це вимагає від системи протидії неперервної адаптації, навчання, оновлення та вдосконалення її аналітичних, технічних та операційних спроможностей, поглиблення можливостей штучного інтелекту для прискорення реагування та розширення джерел.

Подальший технічний розвиток запропонованого автоматизованого механізму має бути спрямований на комплексне підвищення його точності, повноти охоплення інформаційного простору, швидкості виявлення та реагування, глибини аналітичних можливостей та загальної адаптивності до мінливого ландшафту загроз. Це передбачає роботу за кількома ключовими напрямками.

Вдосконалення аналітичних спроможностей на основі ШІ - це пріоритетний напрямок, що включає розширення спектру підтримуваних мов для аналізу багатомовного контенту, що потребуватиме розробки або адаптації відповідних лінгвістичних моделей. Критично важливою є інтеграція досконаліших моделей машинного навчання (МН) та штучного інтелекту (ШІ), зокрема перехід до трансформерних архітектур (таких як, GPT-подібні моделі) для значно точнішого виявлення фейків, пропаганди, мови ворожнечі, аналізу тональності та ідентифікації складних маніпулятивних технік.

Слід розвивати моделі для виявлення прихованих наративів, аналізу специфічних пропагандистських прийомів та ідентифікації скоординованої неавтентичної поведінки (СІВ), можливо, з використанням графових нейронних мереж для аналізу складних взаємозв'язків.

Впровадження технологій автоматичного та напівавтоматичного фактчекінгу шляхом зіставлення тверджень з верифікованими базами даних та авторитетними джерелами є ще одним важливим кроком. Необхідно також забезпечити розширений аналіз мультимедійного контенту, вдосконалюючи модулі для виявлення сучасних діпфейків (відео та аудіо), маніпуляцій із зображеннями.

Нарешті, розвиток прогнозного моделювання та предиктивної аналітики на базі ШІ дозволить прогнозувати потенційне поширення та вплив інформаційних загроз, ідентифікувати вразливі групи населення та передбачати ймовірні сценарії розвитку інформаційних кампаній для більш ефективного превентивного реагування.

Для забезпечення всеосяжного моніторингу необхідно постійно розширювати перелік джерел. Це включає інтеграцію з новими та нішевими онлайн-платформами, які можуть використовуватися для поширення дезінформації (спеціалізовані соціальні мережі, нові месенджери, відеохостинги, ігрові платформи, тематичні форуми), для виявлення підготовки масштабних інформаційних атак або координації деструктивних дій.

Важливим є постійне покращення та адаптація механізмів збору даних, зокрема технологій обходу захисту від скрейпінгу (використання ротації IP-адрес, розподілених мереж проксі-серверів, headless-браузерів, що емулюють поведінку людини), оскільки власники веб-ресурсів також вдосконалюють свої методи захисту.

Ключовою метою є максимальна автоматизація рутинних процесів для вивільнення людських ресурсів для складніших аналітичних завдань. Це передбачає автоматизоване формування, структурування та оновлення бази знань про загрози, суб’єктів, кампанії, наративи та методи протидії на основі оброблюваних даних та результатів аналізу ШІ, включаючи автоматичне виявлення зв'язків та патернів. Важливою є інтеграція елементів систем підтримки прийняття рішень (СППР), які на основі аналізу даних, проведеного ШІ, пропонуватимуть операторам оптимальні варіанти реагування,

оцінюватимуть потенційну ефективність різних контрзаходів та допоможуть у плануванні проактивних інформаційних кампаній.

Поглиблене впровадження практик MLOps (Machine Learning Operations) забезпечить безперервний моніторинг ефективності моделей машинного навчання, їх автоматичне або напівавтоматичне донавчання на нових даних (включаючи дані, верифіковані експертами за принципом HITL), версіонування моделей та даних, а також автоматизоване розгортання оновлених моделей, що є критично важливим для підтримки адаптивності системи до постійно мінливого ландшафту загроз.

Зі зростанням обсягів даних та кількості користувачів, система має зберігати високу продуктивність. Це вимагає постійної оптимізації алгоритмів, запитів до баз даних, архітектурних рішень та раціонального використання обчислювальних ресурсів. Необхідне посилення заходів кібербезпеки для захисту самої системи та даних, що в ній обробляються, включаючи регулярний аудит безпеки, тестування на проникнення, впровадження новітніх технологій захисту від несанкціонованого доступу, витоків даних та атак на моделі ШІ (наприклад, data poisoning).

Успішне впровадження та ефективне, стале використання розробленого автоматизованого механізму, наприклад, на базі ЦПД, вимагає не лише технічної досконалості системи, але й комплексного, стратегічно виваженого підходу, що охоплює організаційні, правові, фінансові, технічні та кадрові аспекти.

Етапи та умови впровадження мають бути ретельно сплановані. Початковим кроком є реалізація пілотного проекту та застосування ітеративного підходу до впровадження. Це передбачає запуск механізму в обмеженому масштабі (Minimum Viable Product) на базі визначених підрозділів для тестування в реальних умовах, збору зворотного зв'язку та доопрацювання. Подальше розгортання має відбуватися поетапно, з поступовим розширенням функціоналу та кількості користувачів.

Важливою передумовою є розробка та адаптація нормативно-правової бази, що чітко регламентуватиме використання автоматизованих систем

моніторингу, збору та обробки даних (з неухильним дотриманням законодавства про захист персональних даних та права на приватність), а також визначатиме порядок міжвідомчої взаємодії та реагування на інформаційні загрози.

Критично важливою є підготовка, безперервне навчання та утримання кваліфікованого персоналу: аналітиків, операторів моніторингу, спеціалістів з даних (data scientists), інженерів машинного навчання (ML engineers), DevOps-інженерів, експертів з кібербезпеки та спеціалістів зі стратегічних комунікацій. Навчальні програми мають охоплювати як технічні аспекти роботи з системою, так і методологію аналізу інформаційних загроз, основи OSINT, психологію впливу, правові та етичні норми.

У процесі експлуатації неминуче виникатимуть можливі виклики, до яких необхідно динамічно адаптовуватись. Технологічні виклики, пов'язані зі швидкою еволюцією методів дезінформації (наприклад, розвиток генеративного ШІ, що ускладнює детекцію діпфейків), вимагають постійного науково-технічного моніторингу. Ресурсні обмеження (фінансування, апаратне забезпечення) потребують чіткого економічного обґрунтування потреби в належному та сталому фінансуванні, пошуку можливостей залучення міжнародної технічної допомоги та оптимізації витрат. Кадровий голод, особливо на спеціалістів з ШІ та аналізу даних, може бути подоланий через державні програми підготовки, тісну співпрацю з університетами. Правові та етичні питання, пов'язані з моніторингом інформаційного простору, збором та обробкою великих масивів даних (включаючи потенційно персональні), можливістю помилкових спрацювань системи та звинувачень у цензурі, вимагають розробки чіткого нормативного регулювання, забезпечення прозорості (в допустимих межах).

Очікуваний ефект від впровадження механізму є багатогранним та стратегічно важливим для держави. Це, насамперед, підвищення оперативності виявлення та швидкості реагування на виклики інформаційної війни. Також очікується поглиблення розуміння тактик супротивника. Важливим є зменшення впливу людського фактору при первинній оцінці інформаційних потоків.

Система дозволить накопичувати та систематизувати унікальну базу знань про загрози та методи протидії, що підвищить інституційну спроможність держави. Кінцевою метою є зміцнення національної інформаційної безпеки та підвищення стійкості українського суспільства до деструктивних інформаційно-психологічних впливів.

Висновки до розділу 3

У третьому розділі кваліфікаційної роботи було розроблено концептуальну модель автоматизованого механізму протидії інформаційній війні, обґрунтовано його необхідність, визначено ключові вимоги, архітектуру, функціональні компоненти та перспективи розвитку й інтеграції.

У ході виконання цього завдання було обґрунтовано доцільність впровадження запропонованого механізму в діяльність існуючих державних структур, таких як Центр протидії дезінформації, для посилення їхніх спроможностей, та сформульовано комплексні функціональні й нефункціональні вимоги до автоматизованої системи, що базується на технологіях штучного інтелекту.

Представлено загальну модель функціонування запропонованого механізму, що описує послідовні етапи обробки інформації: від збору даних з різноманітних джерел, їх попередньої обробки та централізованого зберігання до аналізу за допомогою Аналітичного ядра, управління виявленими інцидентами, оперативного сповіщення та надання користувачам інструментів візуалізації й взаємодії.

Здійснено проектування модульної архітектури механізму, включаючи опис призначення, ключових функцій, практичних аспектів реалізації, потенційних технологічних рішень та можливих ризиків для кожного з основних компонентів.

Практична перевірка теоретичних положень роботи здійснювалася шляхом розробки та тестування програмного прототипу компонента системи, що

інтегрує функції збору даних з веб-сайтів та відеохостингу YouTube, їх попередньої обробки, базової класифікації контенту за ключовими словами та автоматизованого сповіщення через Telegram-бот. Результати тестування підтвердили працездатність запропонованих підходів та визначили напрямки для подальшого вдосконалення аналітичних можливостей.

Таким чином, у третьому розділі представлено цілісну концепцію створення та функціонування автоматизованого механізму протидії інформаційній війні, що закладає основу для його потенційної реалізації та подальшого розвитку.

ВИСНОВКИ

У кваліфікаційній роботі проведено комплексне дослідження проблеми протидії інформаційній війні та розроблено концептуальну модель автоматизованого механізму, спрямованого на підвищення ефективності такої протидії.

1) У ході виконання першого завдання було проаналізовано теоретико-правові засади та сучасний стан проблеми інформаційної війни. Досліджено сутність цього явища, класифіковано основні загрози, розглянуто роль соціальних мереж як ключових інструментів поширення інформаційних впливів та проаналізовано особливості чинного нормативно-правового регулювання в Україні, виявивши його сильні сторони та існуючі прогалини.

2) В рамках другого завдання було досліджено існуючі методи та практичний досвід протидії інформаційним загрозам. Проаналізовано національний та міжнародний досвід, механізми виявлення дезінформації, що застосовуються провідними онлайн-платформами, та проведено оцінку ефективності поточних рішень, що дозволило ідентифікувати ключові виклики та вразливості, зокрема високий рівень поширення деструктивного контенту в Telegram.

3) Третє завдання роботи полягало у розробці концептуальної моделі автоматизованого механізму протидії інформаційним загрозам. Було обґрунтовано необхідність впровадження такого механізму для посилення існуючих державних структур, сформульовано функціональні та нефункціональні вимоги до системи, запропоновано комплексну модульну архітектуру та деталізовано функції її компонентів. Для практичної апробації було розроблено та протестовано програмний прототип. На завершення, визначено основні перспективи подальшого технічного розвитку та стратегічної інтеграції розробленого механізму.

Таким чином, усі поставлені завдання кваліфікаційної роботи виконано, мета роботи досягнута.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Дубов Д. В. Кібербезпека: соціотехнічний аспект [Електронний ресурс] / Д. В. Дубов ; Національний інститут стратегічних досліджень. Режим доступу: https://niss.gov.ua/publikatsiyi

2. Kalyuzhnyi, R. Strategic communications of Ukraine during the Russian-Ukrainian war (based on the materials of the Telegram channel Insider Ukraine) [Electronic resource] / R. Kalyuzhnyi, O. Kanishcheva, V. Kauk, Y. Kyselova // Estudios de Psicología. 2022. Vol. 43, No. 3. Access: https://www.redalyc.org/journal/6948/694873801003/html/

3. Про медіа : Закон України від 13.12.2022 р. № 2849-IX (редакція від 31.03.2023) [Електронний ресурс] / Верховна Рада України. Режим доступу: https://zakon.rada.gov.ua/laws/show/2849-20#Text

4. Про основні засади забезпечення кібербезпеки України : Закон України від 05.10.2017 р. № 2163-VIII [Електронний ресурс] / Верховна Рада України. Режим доступу: https://zakon.rada.gov.ua/laws/show/2163-19#Text

5. Про національну безпеку України : Закон України від 21.06.2018 р. № 2469-VIII [Електронний ресурс] / Верховна Рада України. Режим доступу: https://zakon.rada.gov.ua/laws/show/2469-VIII#Text

6. Кремлівська гідра: 300+ телеграм-каналів, які отруюють український інфопростір [Електронний ресурс] / Детектор медіа. 14.12.2022. Режим доступу: https://detector.media/monitorynh-internetu/article/205954/2022-12-14-kremlivska-gidra-300-telegram-kanaliv-yaki-otruyuyut-ukrainskyy-infoprostir/

7. Russo-Ukrainian war disinformation detection in suspicious Telegram channels [Electronic resource] / S. Kovalchuk [et al.]. ResearchGate. March 2025. Access: https://www.researchgate.net/publication/389713687_Russo-Ukrainian_war_disinformation_detection_in_suspicious_Telegram_channels

8. Центр протидії дезінформації [Електронний ресурс] / Рада національної безпеки і оборони України. Режим доступу: https://cpd.gov.ua/

9. StopFake.org: боротьба з фейками про Україну [Електронний ресурс]. Режим доступу: https://www.stopfake.org/uk/

10. Підхід НАТО у галузі боротьби з інформаційними загрозами [Electronic resource] / NATO. 2024. Access: https://www.nato.int/cps/uk/natohq/topics_219728.htm

11. Confessore, N. Twitter Purges Its Fake Followers To Restore the Power of Influence [Electronic resource] / Nicholas Confessore, Gabriel J.X. Dance // The New York Times. 2018. 12 липня. Access: https://www.nytimes.com/2018/07/11/technology/twitter-fake-followers.html

12. The NATO Strategic Communications Centre of Excellence [Electronic resource] / NATO StratCom COE. 2025. Access: https://stratcomcoe.org/about_us/about-nato-stratcom-coe/5

13. Decoding the Information Environment: NATO's Strategic Communications Centre of Excellence [Electronic resource] / NATO Allied Command Transformation. 2024. Access: https://www.act.nato.int/article/stratcom-coe-2024/

14. Information Environment Simulation Platform «InfoRange» [Electronic resource] / NATO StratCom COE. 2025. Access: https://stratcomcoe.org/projects/information-environment-simulation-platform-inforange/3

15. Locked Shields [Electronic resource] / NATO CCDCOE. 2025. Access: https://ccdcoe.org/locked-shields/

16. WMGIC x NATO Countering Disinformation Challenge 2024 [Electronic resource] / NATO CCDCOE. 2024. Access: https://ccdcoe.org/library/publications/wmgic-x-nato-countering-disinformation-challenge-2024/

17. Protect the Future [Electronic resource] / NATO. 2025. Access: https://www.nato.int/protect-the-future/

18. WMGIC x NATO Countering Disinformation Challenge 2024: Publication of Winning Solutions [Electronic resource] / NATO CCDCOE. 2025. Access:

https://ccdcoe.org/news/2025/wmgic-x-nato-countering-disinformation-challenge-2024

19. Shu, K. Detecting Fake News on Social Media / K. Shu, H. Liu. Morgan & Claypool Publishers, 2019. 130 p. (Synthesis Lectures on Data Mining and Knowledge Discovery).

20. How Our Fact-Checking Program Works [Electronic resource] / Meta. Access: https://transparency.meta.com/features/how-fact-checking-works

21. Testing Begins for Community Notes on Facebook, Instagram and Threads [Electronic resource] / Meta. March 2025. Access: https://about.fb.com/news/2025/03/testing-begins-community-notes-facebook-instagram-threads/

22. New rules for labeling and removing synthetic and manipulated media on Twitter [Electronic resource] / LSU Manship School Mass Communication. Feb 2020. Access: https://faculty.lsu.edu/fakenews/elections/new-rules.php

23. The Four Rs of Responsibility, Part 1: Removing harmful content [Electronic resource] / YouTube Blog. 2019. Access: https://blog.youtube/inside-youtube/the-four-rs-of-responsibility-remove/

24. Telegram Moderation Overview [Electronic resource] / Telegram. 2025. Access: https://telegram.org/moderation

25. Permana, A. A. Microservices Software Architecture: A Review [Electronic resource] / A. A. Permana, I. P. A. E. Pratama // Jurnal Sistem dan Informatika (JSON). 2021. Vol. 16, No. 2. P. 105-111. Access: https://doi.org/10.30865/json.v5i2.3664

26. Human-in-the-Loop Machine Learning: What It Is and How It Works [Electronic resource] / Pareto.ai. Access: https://pareto.ai/blog/human-in-the-loop

27. spaCy: Industrial-Strength Natural Language Processing in Python [Electronic resource] / Explosion AI. Access: https://spacy.io/

28. Blei, D. M. Latent Dirichlet Allocation / D. M. Blei, A. Y. Ng, M. I. Jordan // Journal of Machine Learning Research. 2003. Vol. 3. P. 993-1022.

29. Lee, D. D. Algorithms for Non-negative Matrix Factorization [Electronic resource] / D. D. Lee, H. S. Seung // Advances in Neural Information Processing

Systems 13 (NIPS 2000). 2001. P. 556-562. Access: https://proceedings.neurips.cc/paper/2000/file/f9d1152547c0bde01830b7e8bd60024c-Paper.pdf

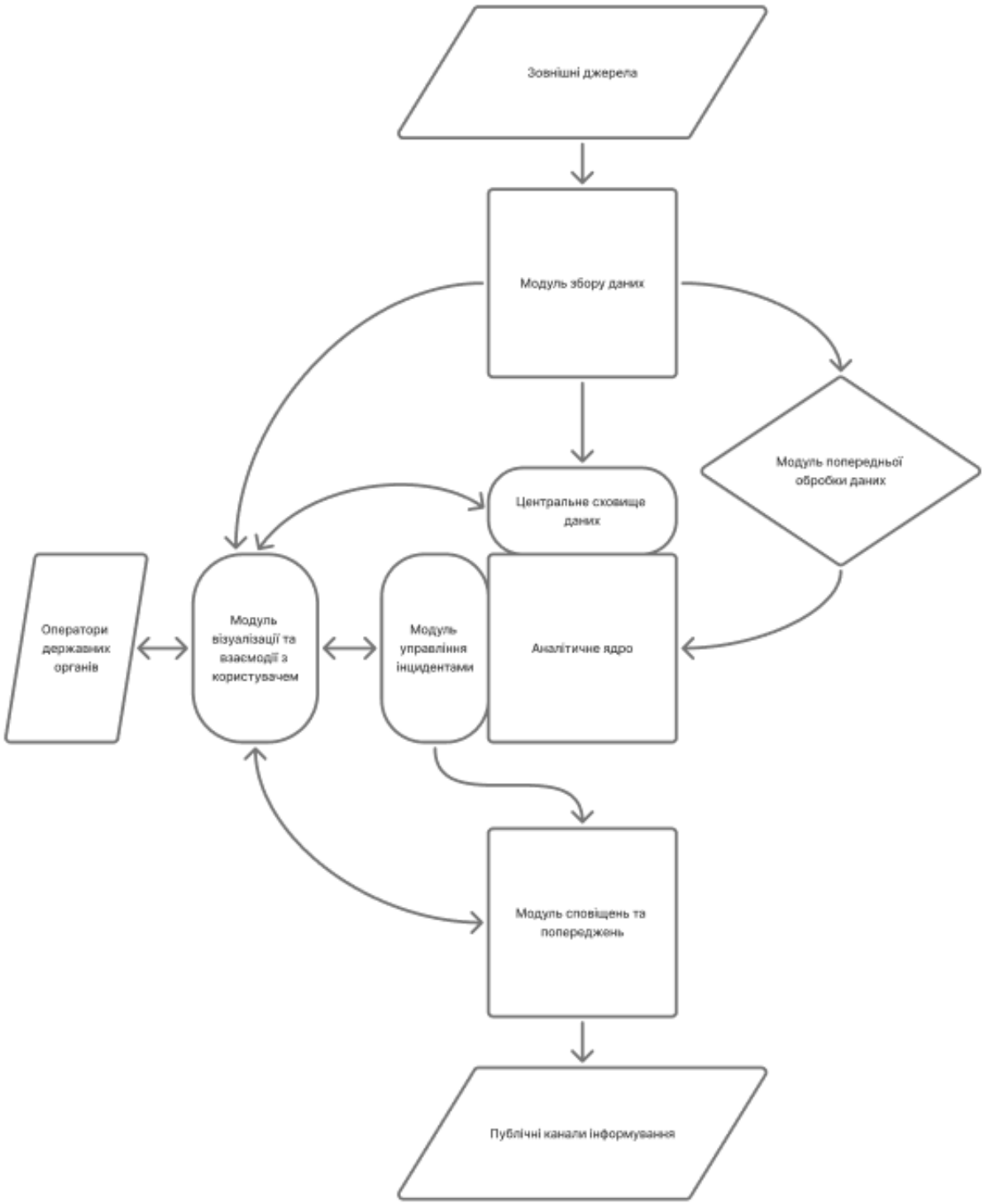
30. Cinelli, M. Coordinated Inauthentic Behavior and Information Spreading on Twitter [Electronic resource] / M. Cinelli, S. Cresci, W. Quattrociocchi // Decision Support Systems (Preprint arXiv:2503.15720). 2025. Access: https://arxiv.org/pdf/2503.15720

31. A Comprehensive Guide on Open Source Intelligence Tools and Techniques [Electronic resource] / Neotas. Access: https://www.neotas.com/osint-tools-and-techniques/

ДОДАТКИ

Додаток А

Загальна блок-схема пропонованої системи



Додаток Б

Код практичної реалізації

```
import requests
from bs4 import BeautifulSoup
import re
import pandas as pd
import time
from googleapiclient.discovery import build
from googleapiclient.errors import HttpError
import json

URL_FILE_PATH = "urls.txt"
KEYWORD_FILE_PATH = "keywords.txt"
CSV_OUTPUT_FILE = "classification_results_ACIW_full.csv"

MANIPULATION_THRESHOLD = 2
REQUEST_DELAY = 1

# --- YouTube Data API
YOUTUBE_API_KEY = "AIzaSyA4UIwUdbXFvP-
GsIdctxJBUI_dpYR04XE"
YOUTUBE_VIDEOS_TO_CHECK_PER_KEYWORD = 1
YOUTUBE_COMMENTS_PER_VIDEO = 2
YOUTUBE_KEYWORDS_TO_SEARCH_LIMIT = 3

NOTIFICATION_BOT_URL = "http://127.0.0.1:5001/send_alert"

# --- МОДУЛЬ ЗАВАНТАЖЕННЯ ДАНИХ З ФАЙЛІВ ---
def load_items_from_file(filepath):
    items = []
    try:
        with open(filepath, 'r', encoding='utf-8') as f:
            for line in f:
                stripped_line = line.strip()
                if stripped_line and not stripped_line.startswith('#'):
                    if filepath == KEYWORD_FILE_PATH:
                        items.append(stripped_line.lower())
                    else:
                        items.append(stripped_line)
        if not items and filepath == URL_FILE_PATH:
          print(f"[!] Попередження: Файл {filepath} порожній або містить лише
коментарі/порожні рядки.")
```

```
        elif not items and filepath == KEYWORD_FILE_PATH:
            print(f"[!] Попередження: Файл {filepath} порожній. Класифікація та пошук на YouTube будуть неефективними.")
        return items
    except FileNotFoundError:
        print(f"[X] Помилка: Файл не знайдено - {filepath}")
        return []
    except Exception as e:
        print(f"[X] Помилка при читанні файлу {filepath}: {e}")
        return []

# --- МОДУЛЬ ЗБОРУ ДАНИХ З ВЕБ-САЙТІВ (WEB SCRAPER) ---
def get_article_text_and_title_from_url(url):
    try:
        headers = {
                'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/96.0.4664.110 Safari/537.36 ACIW_PreviewBot/1.0'
        }
        print(f"   [>] Завантаження URL: {url} (затримка {REQUEST_DELAY} сек)")
        time.sleep(REQUEST_DELAY)
                response = requests.get(url, timeout=25, headers=headers, allow_redirects=True)
        response.raise_for_status()
                    response.encoding = response.apparent_encoding if response.apparent_encoding else 'utf-8'

        soup = BeautifulSoup(response.text, 'html.parser')

        title_text = "Заголовок не знайдено"
        title_tag = soup.find('h1')
        if title_tag:
            title_text = title_tag.get_text(strip=True)
        elif soup.title and soup.title.string:
            title_text = soup.title.string.strip()

        article_text_content = ""
        possible_selectors = [
            {'tag': 'article'},
            {'tag': 'div', 'attrs': {'class': 'article-content'}},
            {'tag': 'div', 'attrs': {'class': 'entry-content'}},
            {'tag': 'div', 'attrs': {'itemprop': 'articleBody'}},
            {'tag': 'div', 'attrs': {'class': 'post-content'}},
            {'tag': 'div', 'attrs': {'class': 'td-post-content'}},
```

```
            {'tag': 'div', 'attrs': {'class': 'content_body'}},
            {'tag': 'div', 'attrs': {'id': 'article_body'}},
            {'tag': 'div', 'attrs': {'class': 'content'}},
            {'tag': 'main'},
        ]

        for selector in possible_selectors:
            article_body = soup.find(selector.get('tag'), selector.get('attrs'))
            if article_body:
                tags_to_remove = ['script', 'style', 'form', 'nav', 'aside', 'footer', 'header',
'iframe', 'noindex']
                    classes_to_remove = ['comments', 'sidebar', 'related-posts', 'social-
share', 'advertisement', 'ad', 'banner']

                for unwanted_tag_name in tags_to_remove:
                    for unwanted_tag in article_body.find_all(unwanted_tag_name):
                        unwanted_tag.decompose()
                for unwanted_class_name in classes_to_remove:
                            for unwanted_tag in article_body.find_all(attrs={'class':
unwanted_class_name}):
                        unwanted_tag.decompose()

                paragraphs = article_body.find_all('p')
                if paragraphs:
                        article_text_content = "\n".join([p.get_text(strip=True) for p in
paragraphs if p.get_text(strip=True)])
                else:
                            article_text_content = article_body.get_text(separator='\n',
strip=True)

                if len(article_text_content.strip()) > 100:
                    break
                else:
                    article_text_content = ""

        if not article_text_content.strip():
            if soup.body:
                for unwanted_tag_name in ['script', 'style', 'form', 'nav', 'aside', 'footer',
'header', 'iframe', 'noindex']:
                    for unwanted_tag in soup.body.find_all(unwanted_tag_name):
                        unwanted_tag.decompose()
                body_text = soup.body.get_text(separator='\n', strip=True)
                 lines = [line for line in body_text.split('\n') if len(line.split()) > 5 and
len(line) > 40]
                article_text_content = "\n".join(lines)
```

```
            else:
                article_text_content = ""

        if not article_text_content.strip():
            print(f"  [!] Не вдалося витягнути основний текст статті для {url}. Можливо, потрібні інші селектори або сайт захищений.")
            return title_text, None

        return title_text, article_text_content

    except requests.exceptions.Timeout:
        print(f"  [X] Помилка: Таймаут запиту ({url})")
        return "Помилка: Таймаут", None
    except requests.exceptions.HTTPError as e:
        print(f"  [X] Помилка HTTP {e.response.status_code} ({url})")
        return f"Помилка HTTP {e.response.status_code}", None
    except requests.exceptions.RequestException as e:
        print(f"  [X] Помилка мережевого запиту ({url}): {e}")
        return "Помилка запиту", None
    except Exception as e:
        print(f"  [!] Загальна помилка при обробці URL ({url}): {e}")
        return "Загальна помилка обробки", None

# --- МОДУЛЬ ЗБОРУ ДАНИХ З YOUTUBE ---
def fetch_youtube_data(api_key, search_query, max_videos_to_fetch=1, max_comments_per_video=2):
    youtube_data_list = []
    if not api_key or api_key == "ВАШ_YOUTUBE_API_KEY":
        print("  [!] Попередження: YOUTUBE_API_KEY не налаштовано. Пропуск YouTube для цього запиту.")
        return []

    try:
        youtube_service_name = "youtube"
        youtube_api_version = "v3"
        youtube = build(youtube_service_name, youtube_api_version, developerKey=api_key)
        print(f"  [>] Пошук відео на YouTube за запитом: '{search_query}' (ліміт: {max_videos_to_fetch} відео)")

        search_request = youtube.search().list(
            q=search_query,
            part='snippet',
            type='video',
            maxResults=max_videos_to_fetch,
```

```
            relevanceLanguage='uk',
            regionCode='UA'
        )
        search_response = search_request.execute()

        video_items = search_response.get('items', [])
        if not video_items:
            print(f"    [i] Відео за запитом '{search_query}' не знайдено на YouTube.")
            return []

        print(f"    [+] Знайдено {len(video_items)} відео для запиту '{search_query}'. Завантаження коментарів (ліміт: {max_comments_per_video} на відео)...")

        for item in video_items:
            video_id = item.get('id', {}).get('videoId')
            video_snippet = item.get('snippet', {})
            video_title = video_snippet.get('title', 'Без назви')
            video_description = video_snippet.get('description', '')
            video_channel_title = video_snippet.get('channelTitle', '')
            video_publish_time = video_snippet.get('publishTime', '')

            print(f"      -> Обробка відео: '{video_title}' (ID: {video_id})")

            video_data_for_analysis = video_title + "\n" + video_description
            if not video_data_for_analysis.strip():
                video_data_for_analysis = "Відео без текстового опису"

            youtube_data_list.append({
                "source_type": "YouTube Video",
                "source_name": video_channel_title,
                "id": video_id,
                "url": f"https://www.youtube.com/watch?v={video_id}",
                "title": video_title,
                "text_content": video_data_for_analysis,
                "author": video_channel_title,
                "created_utc": video_publish_time,
                "score": 0,
                "num_comments_retrieved": 0
            })
            video_entry_index = len(youtube_data_list) - 1

            if max_comments_per_video > 0:
                try:
```

```
                comment_request = youtube.commentThreads().list(
                    part='snippet',
                    videoId=video_id,
                    textFormat='plainText',
                    maxResults=max_comments_per_video,
                    order='relevance'
                )
                comment_response = comment_request.execute()

                comments_count_for_this_video = 0
                for item_comment in comment_response.get('items', []):
                    comment_snippet_outer = item_comment.get('snippet', {})
                                                                top_level_comment =
comment_snippet_outer.get('topLevelComment', {})
                    comment_snippet_inner = top_level_comment.get('snippet', {})

                    comment_text = comment_snippet_inner.get('textDisplay', '')
                                                                comment_author =
comment_snippet_inner.get('authorDisplayName', 'Анонім')
                                                                comment_published_at =
comment_snippet_inner.get('publishedAt', '')
                    comment_like_count = comment_snippet_inner.get('likeCount',
0)
                                                comment_id = top_level_comment.get('id',
item_comment.get('id'))

                    if comment_text:
                        comments_count_for_this_video +=1
                        youtube_data_list.append({
                            "source_type": "YouTube Comment",
                          "source_name": f"Коментар до відео: {video_title[:30]}...",
                            "id": comment_id,
                                                                "url":
f"https://www.youtube.com/watch?v={video_id}&lc={comment_id}",
                                                "title": f"Коментар від {comment_author} до:
{video_title[:40]}...",
                            "text_content": comment_text,
                            "author": comment_author,
                            "created_utc": comment_published_at,
                            "score": comment_like_count,
                            "num_comments_retrieved": 0
                        })

                        if video_entry_index is not None and video_entry_index <
len(youtube_data_list) :
```

```
                    youtube_data_list[video_entry_index]["num_comments_retriev
ed"] = comments_count_for_this_video

                print(f"      [+] Завантажено {comments_count_for_this_video}
коментарів для відео ID: {video_id}")

            except HttpError as e_comment:
                if e_comment.resp.status == 403 and ('commentsDisabled' in
str(e_comment.content).lower()  or  'disabledcomments'  in
str(e_comment.content).lower()):
                    print(f"      [i] Коментарі для відео ID: {video_id} вимкнені.")
                else:
                    print(f"      [X] Помилка при завантаженні коментарів для
відео ID: {video_id}: {e_comment.resp.status} - {e_comment.content}")
            except Exception as e_gen_comment:
                print(f"      [!] Загальна помилка при обробці коментарів для
відео ID: {video_id}: {e_gen_comment}")
        else:
            print(f"      [i] Завантаження коментарів для відео ID: {video_id}
пропущено (max_comments_per_video = 0).")

    except HttpError as e:
        if e.resp.status == 403 and ('quotaExceeded' in str(e.content).lower() or
'dailyLimitExceeded' in str(e.content).lower()):
            print(f"  [X] Помилка YouTube Data API: Денну квоту вичерпано для
запиту '{search_query}'. Спробуйте пізніше або перевірте квоти.")
        else:
            print(f"  [X] Помилка YouTube Data API для запиту '{search_query}':
{e.resp.status} - {e.content}")
    except Exception as e:
        print(f"  [!] Загальна помилка при роботі з YouTube для запиту
'{search_query}': {e}")

    return youtube_data_list

# --- МОДУЛЬ ПОПЕРЕДНЬОЇ ОБРОБКИ ТЕКСТУ ---
def clean_text_for_keyword_search(text_content):
    if not text_content:
        return ""
    processed_text = str(text_content).lower()
    processed_text = re.sub(r'[^\w\s]', ' ', processed_text, flags=re.UNICODE)
    processed_text = re.sub(r'\s+', ' ', processed_text).strip()
    return processed_text

# --- МОДУЛЬ КЛАСИФІКАЦІЇ ---
```

```
def classify_text_by_keywords(text_to_analyze, keyword_list, classification_threshold):
    if not text_to_analyze or not keyword_list:
        return "недостатньо даних для аналізу", 0, []
    unique_found_keywords = set()
    for keyword_or_phrase in keyword_list:
        if keyword_or_phrase in text_to_analyze:
            unique_found_keywords.add(keyword_or_phrase)

    number_of_unique_hits = len(unique_found_keywords)

    if number_of_unique_hits >= classification_threshold:
        classification_result = "[!] Потенційно маніпулятивна"
    else:
        classification_result = "[OK] Нейтральна (або не виявлено маніпуляції)"

    return classification_result, number_of_unique_hits, sorted(list(unique_found_keywords))

# --- Надсилання сповіщення боту ---
def trigger_telegram_alert(message_text):
    """
    Надсилає HTTP POST-запит до Telegram-бота зі сповіщенням.
    """
    try:
        payload = {"message": message_text}
        headers = {'Content-Type': 'application/json'}

        if NOTIFICATION_BOT_URL == "http://localhost:5001/send_alert" and "ВАШ_TELEGRAM_BOT_TOKEN" in open("notification_bot.py", "r", encoding="utf-8").read(): # Примітивна перевірка
            print(f"    [!] Попередження: URL Telegram-бота для сповіщень не налаштований або токен бота не вказано. Сповіщення не буде надіслано.")
            return

        response = requests.post(NOTIFICATION_BOT_URL, data=json.dumps(payload), headers=headers, timeout=10)

        if response.status_code == 200:
            print(f"    [+] Сповіщення про загрозу успішно надіслано Telegram-боту.")
        else:
            print(f"    [X] Помилка надсилання сповіщення Telegram-боту: {response.status_code} - {response.text}")
    except requests.exceptions.ConnectionError:
```

```
        print(f"   [X] Помилка: Не вдалося підключитися до Telegram-бота за адресою {NOTIFICATION_BOT_URL}. Переконайтеся, що бот (notification_bot.py) запущено.")
    except FileNotFoundError:
        print(f"   [!] Попередження: Файл notification_bot.py не знайдено для перевірки токена. Сповіщення може не працювати.")
    except Exception as e:
        print(f"   [X] Невідома помилка при надсиланні сповіщення Telegram-боту: {e}")

# --- ГОЛОВНА ФУНКЦІЯ ПРОГРАМИ ---
def run_analysis_pipeline():
    print("[*] Запуск ACIW_preview (попередній аналіз інформаційних загроз)...")
    print("-" * 70)

    target_urls_list = load_items_from_file(URL_FILE_PATH)
    marker_keywords_list = load_items_from_file(KEYWORD_FILE_PATH)

    youtube_api_configured = YOUTUBE_API_KEY and YOUTUBE_API_KEY != "ВАШ_YOUTUBE_API_KEY"

    if not target_urls_list and not (youtube_api_configured and marker_keywords_list):
        print("[X] Жодне джерело даних (URL-адреси з файлу або YouTube з ключовими словами) не налаштоване або списки цілей порожні.")
        print(f"   Перевірте файл urls.txt та налаштування YOUTUBE_API_KEY і файл keywords.txt.")
        print("[+] Роботу ACIW_preview завершено.")
        return

    if not marker_keywords_list:
        print("[!] Список ключових слів порожній. Класифікація буде неефективною, пошук на YouTube за ключовими словами неможливий.")

    all_results_for_csv = []

    # Обробка URL з файлу
    if target_urls_list:
        print(f"\n--- [*] Початок обробки URL з файлу ({len(target_urls_list)} шт.) ---")
        for index, current_url in enumerate(target_urls_list):
            print(f"\n[>] Обробка URL {index + 1}/{len(target_urls_list)}: {current_url}")
```

```
                                        article_title, raw_article_text = get_article_text_and_title_from_url(current_url)

        current_result_entry = {
            "URL/Джерело": current_url,
             "Заголовок": article_title if article_title else "Не вдалося отримати заголовок",
            "Класифікація": "не оброблено",
            "Рахунок (кількість маркерів)": 0,
            "Знайдені маркери": "немає",
            "Тип джерела": "Веб-сайт"
        }

        if raw_article_text:
            cleaned_text = clean_text_for_keyword_search(raw_article_text)
            if cleaned_text:
                classification_label, keyword_hit_count, list_of_found_keywords = classify_text_by_keywords(
                    cleaned_text,
                    marker_keywords_list,
                    MANIPULATION_THRESHOLD
                )

                print(f"   - Заголовок: {current_result_entry['Заголовок']}")
                if list_of_found_keywords:
                            print(f"     - Ключі ({keyword_hit_count} шт.): {', '.join(list_of_found_keywords)}")
                else:
                    print("   - Ключі: не знайдено")
                        print(f"     - Результат: {classification_label} (Рахунок: {keyword_hit_count})")

                current_result_entry.update({
                    "Класифікація": classification_label,
                    "Рахунок (кількість маркерів)": keyword_hit_count,
                         "Знайдені маркери": ", ".join(list_of_found_keywords) if list_of_found_keywords else "немає"
                })

                if classification_label == "[!] Потенційно маніпулятивна":
                    alert_message = (
                        f"🚨 *Виявлено потенційну інформаційну загрозу!*\n\n"
                        f"*Джерело:* Веб-сайт\n"
                        f"*URL:* {current_url}\n"
                        f"*Заголовок:* {current_result_entry['Заголовок']}\n"
```

```
                        f"*Виявлені маркери:* {current_result_entry['Знайдені маркери']} (Рахунок: {keyword_hit_count})\n"
                        f"_Будь ласка, перевірте для подальшого аналізу._"
                    )
                    trigger_telegram_alert(alert_message)
            else:
                message = "текст статті порожній після очищення"
                print(f"   [!] {message.capitalize()} для {current_url}")
                    current_result_entry["Класифікація"] = f"помилка обробки ({message})"
        else:
            message = "не вдалося отримати або обробити текст статті"
            print(f"   [X] {message.capitalize()} для {current_url}")
             current_result_entry["Класифікація"] = f"помилка ({article_title if article_title and article_title.startswith('Помилка') else message})"

        all_results_for_csv.append(current_result_entry)
        print("-" * 40)
    else:
        print("\n[i] Файл urls.txt порожній або не знайдено. Обробка URL з файлу пропускається.")

    # Обробка даних з YouTube
    if youtube_api_configured and marker_keywords_list:
        print(f"\n--- [*] Початок обробки даних з YouTube ---")
                                        keywords_for_youtube_search = marker_keywords_list[:YOUTUBE_KEYWORDS_TO_SEARCH_LIMIT]
        print(f"[i] Буде використано перші {len(keywords_for_youtube_search)} ключових слів для пошуку на YouTube: {', '.join(keywords_for_youtube_search)}")

        for search_query_yt in keywords_for_youtube_search:
                    youtube_items = fetch_youtube_data(YOUTUBE_API_KEY, search_query_yt,
                                    max_videos_to_fetch=YOUTUBE_VIDEOS_TO_CHECK_PER_KEYWORD,
                                    max_comments_per_video=YOUTUBE_COMMENTS_PER_VIDEO)
            if youtube_items:
                for item_index, item in enumerate(youtube_items):
                    item_title = item.get('title', 'Без назви')
                    item_url = item.get('url', item.get('id', 'URL не знайдено'))
                    item_type = item.get("source_type", "YouTube Data")
                            print(f"\n[>] Обробка YouTube {item_type} (запит: '{search_query_yt}'): {item_title[:60]}...")
```

```
                raw_text = item.get("text_content")

                current_result_entry = {
                    "URL/Джерело": item_url,
                    "Заголовок": f"{item_title} (Пошук: {search_query_yt})",
                    "Класифікація": "не оброблено",
                    "Рахунок (кількість маркерів)": 0,
                    "Знайдені маркери": "немає",
                    "Тип джерела": item_type
                }

                if raw_text:
                    cleaned_text = clean_text_for_keyword_search(raw_text)
                    if cleaned_text:
                                        classification_label, keyword_hit_count,
list_of_found_keywords = classify_text_by_keywords(
                                            cleaned_text, marker_keywords_list,
MANIPULATION_THRESHOLD)

                                        print(f"      - {item_type} (фрагмент):
{raw_text[:100].replace(chr(10), ' ')}...")
                                    if list_of_found_keywords: print(f"      - Ключі
({keyword_hit_count} шт.): {', '.join(list_of_found_keywords)}")
                        else: print("    - Ключі: не знайдено")
                            print(f"    - Результат: {classification_label} (Рахунок:
{keyword_hit_count})")
                        current_result_entry.update({
                            "Класифікація": classification_label,
                            "Рахунок (кількість маркерів)": keyword_hit_count,
                             "Знайдені маркери": ", ".join(list_of_found_keywords) if
list_of_found_keywords else "немає"
                        })

                        if classification_label == "[!] Потенційно маніпулятивна":
                            alert_message = (
                                    f"📥 *Виявлено потенційну інформаційну
загрозу!*\n\n"
                                f"*Джерело:* {item_type}\n"
                                f"*URL/ID:* {item_url}\n"
                                f"*Заголовок/Опис:* {item_title}\n"
                                f"*Виявлені маркери:* {current_result_entry['Знайдені
маркери']} (Рахунок: {keyword_hit_count})\n"
                                f"_Будь ласка, перевірте для подальшого аналізу._"
                            )
                            trigger_telegram_alert(alert_message)
```

```
                else:
                    message = f"текст ({item_type}) порожній після очищення"
                    print(f"   [!] {message.capitalize()}")
                    current_result_entry["Класифікація"] = f"помилка обробки ({message})"
            else:
                message = f"не вдалося отримати текст ({item_type})"
                print(f"   [X] {message.capitalize()}")
                current_result_entry["Класифікація"] = f"помилка ({message})"
            all_results_for_csv.append(current_result_entry)
            print("-" * 40)
        # marker1
    elif not marker_keywords_list and youtube_api_configured:
        print("\n[i] Список ключових слів порожній. Пошук на YouTube за ключовими словами неможливий.")
    elif not youtube_api_configured:
        print("\n[!] Попередження: YOUTUBE_API_KEY не налаштовано. Обробка YouTube пропускається.")
        print("    Будь ласка, заповніть YOUTUBE_API_KEY у скрипті.")

    print("\n--- [*] Обробку всіх налаштованих джерел завершено ---")

    if all_results_for_csv:
        try:
            results_dataframe = pd.DataFrame(all_results_for_csv)
            expected_cols = ["URL/Джерело", "Заголовок", "Класифікація",
                             "Рахунок (кількість маркерів)", "Знайдені маркери", "Тип джерела"]
            for col in expected_cols:
                if col not in results_dataframe.columns:
                    results_dataframe[col] = "N/A"

            results_dataframe = results_dataframe[expected_cols]

            results_dataframe.to_csv(CSV_OUTPUT_FILE, index=False, encoding='utf-8-sig')
            print(f"\n[i] Результати аналізу збережено у файл: {CSV_OUTPUT_FILE}")
        except Exception as e:
            print(f"\n[X] Помилка при збереженні CSV файлу ({CSV_OUTPUT_FILE}): {e}")
    else:
        print("\n[i] Немає даних для збереження у CSV файл (можливо, всі джерела повернули помилки або не були налаштовані).")
```

```
    print("\n[+] Роботу ACIW_preview завершено.")

# --- Точка входу в програму ---
if __name__ == '__main__':
    run_analysis_pipeline()
```