**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ**

**«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ**

**імені ІГОРЯ СІКОРСЬКОГО»**

**НАВЧАЛЬНО-НАУКОВИЙ ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ**

**Кафедра інформаційної безпеки**

«На правах рукопису» «До захисту допущено»

УДК \_\_\_\_\_\_\_\_\_\_\_\_\_\_ Завідувач кафедри

\_\_\_\_\_\_\_\_\_ Дмитро ЛАНДЕ

(підпис)

«\_\_\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_2025 р.

**Магістерська дисертація**

**на здобуття ступеня магістра**

**за освітньо-професійною програмою «Системи, технології та математичні методи кібербезпеки»**

**спеціальності 125 «Кібербезпека та захист інформації»**

на тему: Механізм визначення дезінформації в мережі Інтернет

Виконав (-ла): здобувач вищої освіти ІІ курсу, групи \_\_ФБ-42мп\_\_\_\_\_\_\_

(шифр групи

Рябко Дмитро Вячеславович

(прізвище, ім’я, по батькові) (підпис)

Керівник Ткач Володимир Миколайович

(посада, науковий ступінь, вчене звання, прізвище, ім’я, по батькові) (підпис)

Рецензент \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(посада, науковий ступінь, вчене звання, науковий ступінь, прізвище, ім’я, по батькові) (підпис)

Засвідчую, що у цій магістерській дисертації немає запозичень з праць інших авторів без відповідних посилань.

Здобувач вищої освіти \_\_\_\_\_\_\_\_\_\_\_\_\_

(підпис)

Київ – 2025 року

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ**

**«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ**

**імені ІГОРЯ СІКОРСЬКОГО»**

**НАВЧАЛЬНО-НАУКОВИЙ ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ**

Кафедра інформаційної безпеки

Рівень вищої освіти – другий (магістерський)

Спеціальність – 125 «Кібербезпека та захист інформації»

Освітньо-професійна програма «Системи, технології та математичні методи кібербезпеки»

ЗАТВЕРДЖУЮ

Завідувач кафедри

\_\_\_\_\_\_\_\_\_\_ Дмитро ЛАНДЕ

(підпис)

«\_\_\_»\_\_\_\_\_\_\_\_\_\_\_\_\_2025 р.

**ЗАВДАННЯ**

**на магістерську дисертацію здобувачу вищої освіти**

Рябка Дмитра Вячеславовича

(прізвище, ім’я, по батькові)

1. Тема роботи «Механізм визначення дезінформації в мережі Інтернет»

керівник роботи

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Ткач В. М.\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(прізвище, ім’я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від « » \_\_\_\_\_\_\_ 2025 р. №

2. Термін подання здобувачем вищої освіти роботи « » \_\_\_\_\_ 2025 р.

3. Вихідні дані до роботи

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

4. Зміст роботи

5. Перелік ілюстративного матеріалу (із зазначенням плакатів,

презентацій тощо) Презентація до захисту магістерської дисертації.

6. Дата видачі завдання 10 вересня 2025

Календарний план

|  |  |  |  |
| --- | --- | --- | --- |
| № з/п | Назва етапів виконання  магістерської дисертації | Термін виконання етапів магістерської дисертації | Примітка |
|  | Визначення напрямку МД | 11.09.2025 - 18.09.2025 | Виконано |
|  | Визначення теми МД | 20.09.2025 - 29.09.2025 | Виконано |
|  | Збір та аналіз літературних джерел | 02.10.2025 - 17.10.2025 | Виконано |
|  | Формування технічного завдання | 18.10.2025 - 01.11.2025 | Виконано |
|  | Розробка програмного коду | 01.11.2025 - 24.11.2025 | Виконано |
|  | Тестування та удосконалення | 25.11.2025 - 07.12.2025 | Виконано |
|  | Формування графічної частини (презентації) | 08.12.2025 | Виконано |
|  | Підготовка та передзахист МД | 09.12.2025 - 12.12.2025 | Виконано |
|  | Підготовка до захисту МД | 01.12.2025 - 16.12.2025 | Виконано |

Здобувач вищої освіти \_\_\_\_\_\_\_\_\_\_\_\_ Дмитро РЯБКО

(підпис) (**Власне ім’я, ПРІЗВИЩЕ**)

Керівник роботи \_\_\_\_\_\_\_\_\_\_\_\_ Володимир ТКАЧ

(підпис) (**Власне ім’я, ПРІЗВИЩЕ**)

# РЕФЕРАТ

Обсяг роботи 108 сторінок, 18 ілюстрацій, 25 джерел літератури.

**Об’єкт дослідження:** процес ідентифікації та аналізу маніпулятивного контенту та дезінформації в мережі Інтернет.

**Предмет дослідження:** методи та засоби автоматизованої оцінки надійності веб-ресурсів із застосуванням гібридного підходу, що поєднує аналіз метаданих (OSINT) та семантичний аналіз тексту великими мовними моделями (LLM)

**Мета дослідження:** полягає у підвищенні ефективності виявлення дезінформації шляхом розробки програмного механізму, який дозволяє автоматизувати верифікацію джерел, аналіз авторства та перевірку логіко-емоційної складової тексту для розрахунку інтегрального показника довіри.

**Методи дослідження:** У роботі використано методи системного аналізу (для побудови архітектури), розвідки на основі відкритих джерел (OSINT), обробки природної мови (NLP), сентимент-аналізу, веб-скрапінгу та об'єктно-орієнтованого програмування.

**Отримані результати:** Розроблено веб-застосунок (на базі стеку Next.js та Gemini API), який на основі вхідної URL-адреси виконує комплексний аудит публікації. Програма автоматично збирає технічні дані (Whois, SSL, структура DOM) та використовує нейромережеві моделі для глибинного аналізу тексту за п'ятьма групами критеріїв: технічна надійність, репутація автора, верифікація фактів, логічна зв'язність та емоційна маніпулятивність. Реалізовано алгоритм розрахунку зваженого рейтингу довіри (Trust Score), що дозволяє користувачу миттєво оцінити ризик споживання недостовірної інформації.

**Ключові слова:** дезінформація, osint, великі мовні моделі (LLM), NLP, фактчекінг, веб-скрапінг, сентимент-аналіз, інформаційна безпека, Next.js.

# ABSTRACT

Volume of the work is 108 pages, 18 illustrations, and 25 references.

**The object of the research:** the process of identification and analysis of manipulative content and disinformation on the Internet.

**The subject of the research:** methods and tools for the automated assessment of web resource reliability using a hybrid approach that combines metadata analysis (OSINT) and semantic text analysis by large language models (LLMs).

**The aim of the research:** is to increase the efficiency of disinformation detection by developing a software mechanism that automates source verification, authorship analysis, and the validation of logical and emotional text components to calculate an integral trust score.

**Research methods:** The work utilizes methods of system analysis (for architecture design), Open Source Intelligence (OSINT), Natural Language Processing (NLP), sentiment analysis, web scraping, and Object-Oriented Programming (OOP).

**The obtained results:** A web application (based on the Next.js stack and Gemini API) has been developed, which performs a comprehensive audit of a publication based on an input URL. The software automatically collects technical data (Whois, SSL, DOM structure) and utilizes neural network models for deep text analysis across five criteria groups: technical reliability, author reputation, fact verification, logical coherence, and emotional manipulativeness. An algorithm for calculating a weighted Trust Score has been implemented, allowing the user to instantly assess the risk of consuming unreliable information.

**Keywords:** DISINFORMATION, OSINT, LARGE LANGUAGE MODELS (LLM), NLP, FACT-CHECKING, WEB SCRAPING, SENTIMENT ANALYSIS, INFORMATION SECURITY, NEXT.JS.

# 

# ЗМІСТ

[РЕФЕРАТ 4](#_Toc215772744)

[ABSTRACT 4](#_Toc215772745)

[ЗМІСТ 6](#_Toc215772746)

[ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, 7](#_Toc215772747)

[СКОРОЧЕНЬ І ТЕРМІНІВ 7](#_Toc215772748)

[Вступ 9](#_Toc215772749)

# ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ,

# СКОРОЧЕНЬ І ТЕРМІНІВ

**AI** – Artificial Intelligence (штучний інтелект).

**API** – Application Programming Interface (інтерфейс прикладного програмування). **CSS** – Cascading Style Sheets (каскадні таблиці стилів).

**DOM** – Document Object Model (об'єктна модель документа - структура HTML-сторінки, яку ми парсимо).

**GPT** – Generative Pre-trained Transformer (генеративний попередньо навчений трансформер; архітектура LLM).

**HTML** – HyperText Markup Language (мова розмітки гіпертексту).

**HTTP** – HyperText Transfer Protocol (протокол передачі гіпертексту).

**HTTPS** – HyperText Transfer Protocol Secure (захищений протокол передачі гіпертексту).

**JSON** – JavaScript Object Notation (текстовий формат обміну даними).

**LLM** – Large Language Model (велика мовна модель).

**NER** – Named Entity Recognition (розпізнавання іменованих сутностей; метод виділення імен, дат, локацій з тексту).

**NLP** – Natural Language Processing (обробка природної мови).

**OSINT** – Open Source Intelligence (розвідка на основі відкритих джерел).

**REST** – Representational State Transfer (архітектурний стиль взаємодії компонентів розподіленої програми в мережі).

**SDK** – Software Development Kit (набір засобів розробки; у роботі використано Gemini SDK).

**SEO** – Search Engine Optimization (пошукова оптимізація).

**SSL** – Secure Sockets Layer (криптографічний протокол для безпечного зв'язку). **TLS** – Transport Layer Security (протокол захисту транспортного рівня, наступник SSL).

**UI** – User Interface (інтерфейс користувача).

**URL** – Uniform Resource Locator (уніфікований покажчик ресурсу; веб-адреса). **UX** – User Experience (користувацький досвід).

**WHOIS** – Мережевий протокол прикладного рівня для отримання реєстраційних даних про власників доменних імен та IP-адрес.

**Терміни:**

**Клікбейт (Clickbait)** – пейоративний термін, що описує веб-контент, метою якого є отримання доходу від онлайн-реклами за рахунок низької якості або сенсаційних заголовків.

**Тайпосквотінг (Typosquatting)** – форма кіберсквотінгу, що полягає в реєстрації доменних імен, близьких за написанням до адрес популярних сайтів, з розрахунком на помилки користувачів при введенні адреси.

**Фейк (Fake news)** – інформаційне повідомлення, що містить частково або повністю недостовірну інформацію.

**Фактчекінг (Fact-checking)** – перевірка фактичних тверджень у нехудожньому тексті з метою визначення їх правдивості та правильності.

# ВСТУП

**Актуальність роботи.** Сучасний інформаційний простір характеризується безпрецедентною швидкістю поширення даних, що створює сприятливі умови для масових маніпуляцій та дезінформаційних кампаній. Традиційні методи фактчекінгу, які покладаються на ручну перевірку експертами або статичні «чорні списки» доменів, втрачають ефективність проти технологічно вдосконалених фейків, що генеруються автоматично та адаптуються під цільову аудиторію. Використання великих мовних моделей (LLM) та методів розвідки на основі відкритих даних (OSINT) дозволяє не лише автоматизувати процес верифікації джерел, але й проводити глибинний семантичний аналіз контенту, виявляючи приховані логічні хиби та емоційні маніпуляції.

**Мета дослідження** полягає в розробці комплексного програмного механізму визначення дезінформації в мережі Інтернет, що поєднує детермінований аналіз технічних параметрів веб-ресурсів та імовірнісний аналіз текстового контенту за допомогою штучного інтелекту.

**Завдання дослідження:**

1. Провести аналіз існуючих підходів до виявлення фейкових новин та визначити обмеження сучасних автоматизованих систем.
2. Розробити систему критеріїв оцінювання достовірності інформації, згруповану за п'ятьма векторами: технічні показники, репутація автора, верифікація фактів, логічна якість тексту та емоційне забарвлення.
3. Запропонувати та програмно реалізувати гібридну архітектуру веб-застосунку, що інтегрує методи OSINT (Whois, парсинг метаданих) та NLP-сервіси (Generative AI) для розрахунку інтегрального рейтингу довіри.
4. Експериментально перевірити ефективність розробленого механізму на вибірці реальних новинних та дезінформаційних матеріалів, оцінити точність роботи алгоритмів.

**Методи дослідження.** У роботі використано методи системного аналізу, розвідки на основі відкритих джерел (OSINT), обробки природної мови (NLP), сентимент-аналізу, веб-скрапінгу та об’єктно-орієнтованого програмування.

**Наукова новизна** полягає у вдосконаленні методології автоматизованого виявлення дезінформації шляхом створення гібридної моделі оцінювання, яка, на відміну від існуючих аналогів, одночасно враховує технічні атрибути джерела, цифровий слід автора та семантичні патерни маніпуляції в тексті.

**Практичне значення.** Практичне значення роботи полягає у створенні діючого веб-інструменту, який дозволяє користувачам у режимі реального часу оцінювати надійність інформаційних джерел, що сприяє підвищенню рівня медіаграмотності та зменшенню впливу дезінформації на суспільство