Хмельницький національний університет
Факультет інформаційних технологій
Кафедра інженерії програмного забезпечення

# КВАЛІФІКАЦІЙНА РОБОТА

Метод проектування програмних модулів для знаходження неправдивих новин у соціальних мережах
Назва теми

Рівень вищої освіти Другий (магістерський)

Галузь знань 12 «Інформаційні технології»

Спеціальність 121 «Інженерія програмного забезпечення»

Освітня програма Освітньо-професійна програма «Інженерія програмного забезпечення»

Шифр КвРІПЗ.2301116.01.03.ПЗ

Виконав студент 2 курсу, група ІПЗм-23-1 _____ Валентин ВАРУК
Підпис Ініціали, прізвище

Керівник к. техн. наук, доцент _____ Юрій Форкун
Науковий ступінь, звання Підпис Ініціали, прізвище

Нормоконтролер к. пед. наук, доцент _____ Оксана ОНИШКО
Підпис Ініціали, прізвище

**До захисту допускаю:**
Завідувач кафедри інженерії програмного забезпечення _____ Леонід БЕДРАТЮК
Підпис Ініціали, прізвище

2 грудня 2024.

Хмельницький 2024

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

Факультет Інформаційних технологій
Кафедра Інженерії програмного забезпечення
Рівень вищої освіти Другий (магістерський)
Галузь знань 12 «Інформаційні технології»
Спеціальність 121 «Інженерія програмного забезпечення»
Освітня програма Освітньо-професійна програма «Інженерія програмного забезпечення»

ЗАТВЕРДЖУЮ
Завідувач кафедри ІПЗ
Леонід Бедратюк
2.09
2024 р.

# ЗАВДАННЯ
# НА КВАЛІФІКАЦІЙНУ РОБОТУ

Варуку Валентину Костянтиновичу
Прізвище, ім'я, по батькові студента

1. Тема проекту (роботи) Метод проектування програмних модулів для знаходження неправдивих новин у соціальних мережах

Керівник проекту (роботи) канд. техн. наук, доцент Форкун Ю.В.
Прізвище, ім'я, по батькові, науковий ступінь, вчене звання

Затверджена наказом ректора університету від 26.08.2024 р. № 104-КР
2. Строк подання студентом проекту (роботи) на кафедру 02.12.2024 р.
3. Вихідні дані до проекту (роботи) Матеріали науково-дослідної практики
4. Зміст пояснювальної записки (перелік питань, які потрібно розробити)
1 Аналіз предметної області та рішень з програмного забезпечення.
2 Удосконалення методу підтримки якості програмного коду на основі автоматичного тестування
3 Архітектура програмної реалізації
4 Програмна реалізація
5. Перелік графічного матеріалу (із зазначенням обов'язкових креслень)
Презентаційні матеріали (слайди)

6. Консультанти розділів дипломного проекту (роботи)

| Розділ | Прізвище, ініціали та посада консультанта | Підпис, дата | |
|---|---|---|---|
| | | завдання видав | завдання прийн |
| Антиплагіат | Форкун Ю.В., доцент | | |
| Нормоконтроль | Онишко О.Г., доцент | | |

7. Дата видачі завдання «01» вересня 2024 р.

КАЛЕНДАРНИЙ ПЛАН

| Назва етапів (розділів) дипломного проекту (роботи) | Строк виконання етапів проекту (роботи) | Примі |
|---|---|---|
| Вивчення предметної області; формулювання мети та задач дослідження; визначення об'єкта та предмета дослідження | 02 | |
| Робота над розділом 1 кваліфікаційної роботи – вивчення літературних та Інтернет-джерел; аналіз відомих моделей, методів та засобів за темою роботи; | 11 | |
| Робота над розділом 2 кваліфікаційної роботи – розробка моделей, методів та алгоритмів вирішення задачі; висновки до розділу | | |
| 4. Робота над науковими статтями | 3 | |
| Робота над розділом 3 кваліфікаційної роботи – розробка інформаційної технології вирішення задачі | | |
| Робота над розділом 4 кваліфікаційної роботи – програмна реалізація спроектованих рішень, результати експериментів та їх аналіз; | 7 | |
| Попередній захист кваліфікаційної роботи | Листопад (згідно графіка) | |
| Узгодження постановки задачі, отриманих результатів та виснов- ків; оформлення пояснювалюної записки та графічних матеріалів згідно вимог чинних стандартів | | |
| Перевірка роботи на наявність плагіату; нормоконтроль; брошурування пояснювальної записки; підготовка супровідних документів | | |
| Підготовка до захисту дипломної роботи | з 02.12.2024 | |

Студент _____ Валентин ВАРУК
Підпис Ініціали, прізвище

Керівник проекту (роботи) _____ Юрій ФОРКУН
Підпис Ініціали, прізвище

# РЕФЕРАТ

Тема кваліфікаційної роботи: «Метод проектування програмних модулів для знаходження неправдивих новин у соціальних мережах».

Автор роботи: Варук Валентин Костянтинович.

Керівник роботи: Форкун Юрій Вікторович.

Пояснювальна записка: 72 с., 11 рис., 8 таб., 2 дод., 41 джерела.

ФЕЙКОВІ НОВИНИ, НЕПРАВДИВА ІНФОРМАЦІЯ, ШТУЧНИЙ ІНТЕЛЕКТ, ПРОЕКТУВАННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ, ПРОГРАМНІ МОДУЛІ, СОЦІАЛЬНА МЕРЕЖА.

**Мета роботи** – методи та засоби проектування програмних модулів для знаходження неправдивої інформації із використанням методу визначення достовірності даних.

**Об’єктом дослідження є** процеси проектування програмних застосунків для знаходження неправдивої інформації із використанням методу визначення достовірності даних.

**Предметом дослідження є** методи проектування програмних застосунків у сфері кібербезпеки та кібергігієни.

**Наукова новизна:**

– удосконалення методу проектування програмних модулів шляхом введення нових класифікаторів і проектування модулів для ефективного виявлення неправдивої інформації.

Для проведення якісного дослідження, встановлено виконання таких завдань:

– проаналізувати предметну область галузі дослідження, а саме методи, засоби та проектування модулів для знаходження неправдивої інформації;

– проаналізувати існуючі методи проектування модулів для знаходження неправдивих (фейкових новин);

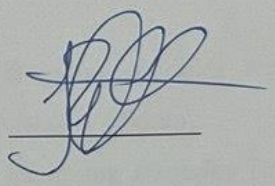
– удосконалити метод проектування програмних модулів шляхом введення нових класифікаторів і проектування модулів для ефективного виявлення неправдивої інформації.

Практичне значення отриманих результатів. У даній кваліфікаційній магістерській роботі пропонується удосконалення методу проектування програмних модулів шляхом введення нових класифікаторів і проектування модуля для ефективного виявлення та відображення неправдивих новин потенційним використанням веб-додатку.

Це дозволить не лише підвищити точність моделі, але й забезпечити більшу стабільність та надійність результатів.

В подальшому результати даного дослідження при розробці програмного забезпечення призначеного для боротьби із розповсюдженням фейкових (неправдивих) новин у різноманітних медійних засобах, зокрема у соціальних мережах.

01.12.24



# ABSTRACT

Master’s thesis: « A method for designing software modules for detecting false news on social media».

Author: Valentyn Varuk.

Head of work: Yurii Forkun.

Master’s thesis consists of: 72 pages of the general text, 11 graphics, 8 tables, 2 supplements, 41 literature sources.

FAKE NEWS, FALSE INFORMATION, ARTIFICIAL INTELLIGENCE, SOFTWARE DESIGN, SOFTWARE MODULES, SOCIAL NETWORK.

The purpose of the study is to design modules for finding false information using the method of determining the reliability of data.

The object of research is the process of designing modules for finding false information using the method of determining the reliability of data.

The subject of research is the methods of designing modules for finding false information.

Scientific novelty:

1. Improving the method of designing software modules by introducing new classifiers and designing a module to effectively detect false information.

To conduct a qualitative study, the following tasks were set:

1. To analyze the subject area of the research field, namely, methods, tools and design of modules for finding false information;

2. To analyze existing methods of designing modules for detecting false (fake news);

3. Improve the method of designing software modules by introducing new classifiers and designing a module to effectively detect false information.

Practical significance of the results. This master's thesis proposes to improve the method of designing software modules by introducing new classifiers and designing a module for the effective detection and display of false news by the potential use of a web

application. This will not only improve the accuracy of the model, but also ensure greater stability and reliability of the results.

In the future, the results of this study will be used to develop software designed to combat the spread of fake news in various media, including social networks.

01.12.24

**ЗМІСТ**

# ВСТУП

Неправдиві (фейкові) новини завжди були критичною і складною проблемою в інформаційному середовищі. Поширення неправдивих новин викликає серйозне занепокоєння, особливо у сфері медичної інформації, що може мати небезпечні та потенційно смертельні наслідки. З огляду на цунамі дезінформації в Інтернеті вкрай важливо боротися з фейковими медичними новинами.

Фейкова або неправдива інформація, в тому числі і новини постійно були критичною і складною проблемою в інформаційному медійному просторі. Поширення неправдивих новин викликає хвилювання та глибоке занепокоєння, особливо у сфері медичної інформації, що може мати небезпечні та потенційно смертельні наслідки.

Також неправдиві новини в соціальних мережах можуть мати непоправно важкі наслідки також в сучасних умовах ведення бойових дій на території України, адже групи та канали створюються із величезною швидкістю, а інформація в них розповсюджується ще швидше.

З огляду на величезні пули дезінформації в мережі Інтернеті вкрай важливо боротися з фейковими новинами, в тому числі медичними та пов’язаними із військовими подіями на території України.

Саме тому використання методів проектування програмних модулів, які б допомогали виявити фейкові новини, в тому числі й медичного та військового (воєнного) спрямування є актуальною проблемою, особливо в умовах воєнних дій на території України, для проведення дослідження.

На сьогоднішній день існують методи побудови таких програмних модулів для виявлення неправдивих новин, які використовують одномодальні ознаки, тобто аналізують лише один тип даних, наприклад, текстову інформацію. Однак, при об'єднанні різних модальностей, таких як текст, зображення та метадані, ефективність цих одномодальних методів може знижуватися. Це зумовлено тим, що ознаки, релевантні для однієї модальності, можуть не бути інформативними або

навіть вводити в оману при аналізі комбінованих даних. У статті [37] автори пропонують підхід, який враховує текстові, візуальні та соціальні ознаки для покращення точності виявлення неправдивих новин. Вони зазначають, що використання лише одномодальних ознак може бути недостатнім, оскільки сучасні фейкові новини часто містять комбінований контент, ігнорування якого може призвести до помилкових висновків.

Таким чином, для підвищення достовірності виявлення неправдивих новин рекомендується застосовувати мультимодальні методи при проектуванні ПЗ, які здатні інтегрувати та аналізувати різні типи даних одночасно.

Загалом, у таких підходах інформація подається у вигляді простого тексту, тому початкові методи переважно виявляють неправдиву інформацію шляхом вилучення текстових ознак з текстового контенту, або вилучення відповідних однонаправлених (одномодальних) ознак з інших однонаправлених даних.

З розвитком соціальних мереж форма інформації змінилася з простого тексту на мультимедійну. Більшість існуючої інформації знаходиться у формі мультимедіа. У той же час поєднання особливостей різних модальностей може ефективно покращити продуктивність моделі. Тому останні дослідження в галузі виявлення неправдивої інформації загалом побудовані із використанням мультимодальних методів.

**Мета роботи** – методи та засоби проектування програмних модулів для знаходження неправдивої інформації із використанням методу визначення достовірності даних.

**Об’єктом дослідження є** процеси проектування програмних застосунків для знаходження неправдивої інформації із використанням методу визначення достовірності даних.

**Предметом дослідження є** методи проектування програмних застосунків у сфері кібербезпеки та кібергігієни.

**Наукова новизна:**

– удосконалення методу проектування програмних модулів шляхом введення нових класифікаторів і проектування модулів для ефективного виявлення неправдивої інформації.

Для проведення якісного дослідження, встановлено виконання таких завдань:

– проаналізувати предметну область галузі дослідження, а саме методи, засоби та проектування модулів для знаходження неправдивої інформації;

– проаналізувати існуючі методи проектування модулів для знаходження неправдивих (фейкових новин);

– удосконалити метод проектування програмних модулів шляхом введення нових класифікаторів і проектування модулів для ефективного виявлення неправдивої інформації.

Практичне значення отриманих результатів. У даній кваліфікаційній магістерській роботі пропонується удосконалення методу проектування програмних модулів шляхом введення нових класифікаторів і проектування модуля для ефективного виявлення та відображення неправдивих новин потенційним використанням веб-додатку. Даний метод дозволить забезпечити більшу стабільність та надійність результатів.

В подальшому результати даного дослідження при розробці програмного забезпечення призначеного для боротьби із розповсюдженням фейкових (неправдивих) новин у різноманітних медійних засобах, зокрема у соціальних мережах.

Відповідно до теми кваліфікаційної роботи опубліковані тези «Аналіз програмних методів та засобів виявлення фейкових новин» на конференції «Актуальні проблеми комп’ютерних наук АПКН-2024».

# 1 ТЕОРЕТИЧНИЙ ВИКЛАД ДОСЛІДЖУВАНОЇ ПРОБЛЕМИ

## 1.1 Аналіз предметної області

У сучасному світі соціальні мережі стали головним джерелом інформації для мільярдів людей. Разом із цим зростає загроза поширення неправдивих новин (fake news), які впливають на громадську думку, сприяють дезінформації та навіть політичній дестабілізації. Традиційні методи розробки програмного забезпечення перевірки фактів не здатні оперативно обробляти величезний обсяг контенту, що публікується щосекунди. Тому необхідно розробляти програмні модулі, здатні автоматично виявляти неправдиві новини, інтегруючи сучасні технології обробки даних і машинного навчання.

Метод проектування програмних модулів для виявлення неправдивих новин ґрунтується на кількох ключових принципах, таких як: мультимодальний аналіз, застосування штучного інтелекту, інтеграція з API, модульна архітектура.

Мультимодальний аналіз при проектуванні такого програмного забезпечення передбачає використання даних із різних джерел. Це може бути текст, зображення, відео та соціальні сигнали для підвищення точності ідентифікації.

– текстовий аналіз передбачає оцінку стилю написання, тональності та перевірку фактів у відповідних базах даних;

– аналіз зображень дозволяє виявляти маніпуляції чи проводити порівняння із джерелами оригінальних зображень;

– соціальні сигнали аналізуються для виявлення шаблонів поширення новин, таких як активність ботів або нетипові сплески поширення.

Цей підхід забезпечує комплексний аналіз, що дозволяє підвищити точність ідентифікації неправдивої інформації та мінімізувати ризики поширення маніпулятивного контенту

На рисунку 1.1 демонструються різні аспекти мультимодального аналізу для виявлення фейків.

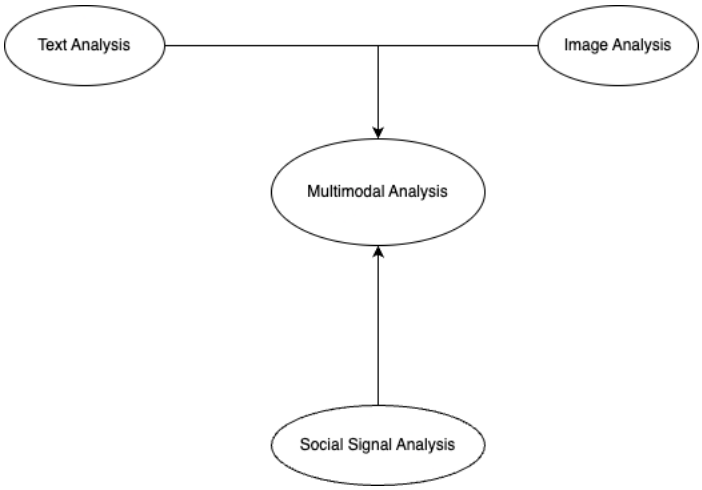



Рисунок 1.1 – Графік мультимодального аналізу

Застосування штучного інтелекту дозволяє проводити текстовий аналіз за допомогою моделей глибокого навчання, таких як BERT або RoBERTa. Нейронні мережі використовуються для обробки зображень, наприклад ResNet або EfficientNet. Графові нейронні мережі допомагають аналізувати взаємодії в соціальних мережах.

Інтеграція з API соціальних платформ забезпечує отримання даних у реальному часі для подальшого аналізу.

Модульна архітектура складається з кількох компонентів, таких як модулі збору даних, попередньої обробки, класифікації та звітності. Кожен із них виконує специфічну задачу для досягнення загальної мети. На рисунку 1.2, продемонстровано блок-схему модульної архітектури.

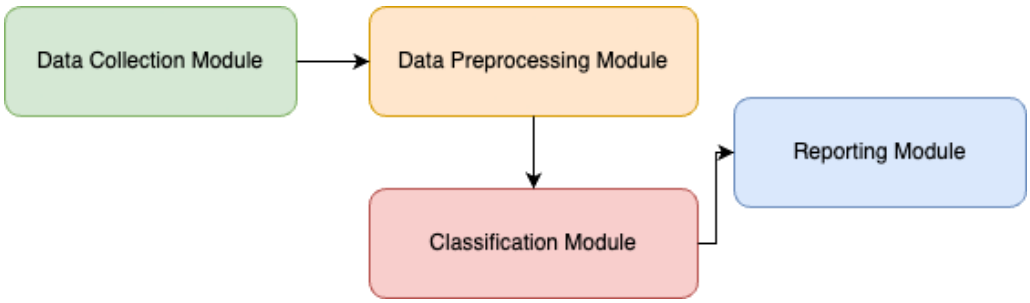



Рисунок 1.2 – Блок-схема модульної архітектури

Розробка модулів для виявлення неправдивих новин має стратегічне значення для забезпечення достовірності інформаційного простору. Основною метою таких модулів є протидія поширенню фейкових новин, які здатні завдати шкоди не лише на індивідуальному, а й на суспільному рівні. Автоматизовані системи виявлення дозволяють ідентифікувати неправдиву інформацію на ранніх етапах її поширення, що суттєво зменшує потенційний вплив фейків на громадську думку.

Рання ідентифікація неправдивих матеріалів мінімізує їхній негативний вплив, знижуючи ймовірність того, що вони встигнуть поширитися та викликати серйозні наслідки. Зокрема, це стосується таких сфер, як виборчі процеси, громадська безпека, соціальна стабільність, економічні ринки тощо. Швидке реагування на дезінформацію дозволяє суспільству отримувати більш точну інформацію і приймати обґрунтовані рішення.

Автоматизація перевірки фактів є ключовим інструментом для підтримки фактчекінгових ініціатив. Завдяки таким модулям журналісти та дослідники отримують можливість швидше і точніше перевіряти інформацію, що знижує навантаження на їхню роботу. Це підвищує якість журналістських матеріалів і забезпечує громадськість достовірними даними.

Крім того, системи автоматизованої перевірки сприяють забезпеченню інформаційної безпеки. У сучасному світі дезінформація може стати загрозою для політичної, економічної чи соціальної стабільності, впливаючи на прийняття рішень як на рівні окремих осіб, так і на рівні державних структур. Модулі для виявлення фейкових новин зменшують ці ризики, захищаючи громадян від маніпуляцій і сприяючи формуванню більш надійного інформаційного середовища.

Деякі існуючі рішення можуть бути адаптовані для цієї задачі. Наприклад, HateDetect аналізує токсичний контент і може бути використаний для виявлення фейкових новин. FakeFlow аналізує поширення неправдивої інформації у соціальних мережах. Модулі на основі BERT дозволяють класифікувати тексти за

допомогою готових моделей машинного навчання. Інструменти, такі як Google Fact-Check Tools, пропонують API для перевірки фактів.

## 1.2 Аналіз існуючих рішень

У сучасному цифровому середовищі соціальні мережі відіграють важливу роль у поширенні інформації, стаючи головним джерелом новин для мільярдів людей по всьому світу. Однак поряд із позитивними аспектами, такими як швидкість обміну даними та доступність інформації, з'являється серйозна загроза, поширення фейкових новин. Цей феномен має значний вплив на суспільство, оскільки неправдива інформація може викликати дезінформацію, маніпулювати громадською думкою, сприяти соціальній напрузі та навіть політичній дестабілізації.

Сучасні виклики, пов'язані з фейковими новинами, стимулювали активні дослідження у сфері їх автоматичного виявлення. Розробники й науковці зосереджують увагу на створенні інноваційних методів, які використовують штучний інтелект, аналіз великих даних та алгоритми машинного навчання. Основна мета таких досліджень – забезпечити швидке виявлення неправдивої інформації, зменшити її поширення та мінімізувати негативний вплив на суспільство. Застосування новітніх технологій дозволяє автоматизувати процес аналізу інформації, враховуючи різноманітні аспекти для визначення джерел і характеру фейкових новин, що сприяє створенню довірливого інформаційного простору, де кожен користувач має доступ до перевіреної та достовірної інформації.

Далі буде розглянуто основні підходи, див. таблицю 1.1, до вирішення цієї проблеми, їхні переваги та недоліки, такі як лінгвістичні методи, методи які засновані на знаннях, соціально-контекстуальні методи, гібридні методи.

На рисунку 1.3 продемонстровано основні підходи, для виявлення фейкових новин.

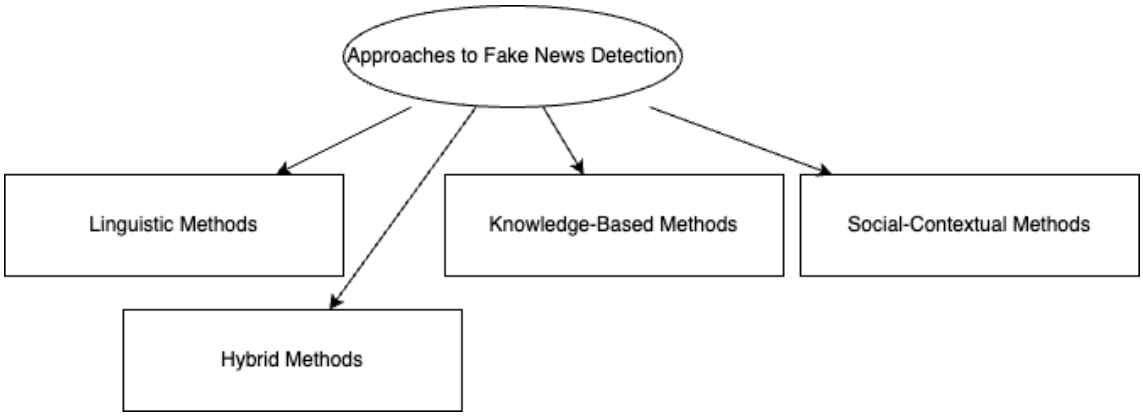



Рисунок 1.3 – Підходи для виявлення фейкових новин

Лінгвістичні методи [38] аналізують текстові характеристики новинних статей для оцінки їх достовірності. Вони базуються на припущенні, що фейкові новини мають специфічні мовні особливості, такі як суб'єктивний тон, використання емоційно забарвлених слів, перебільшення та інші риторичні прийоми. Такі методи дозволяють виявляти маніпулятивний характер тексту шляхом аналізу лексичних і синтаксичних особливостей повідомлення.

Переваги:

– дозволяють виявляти фейкові новини без необхідності додаткових даних про джерело чи поширення;

– можуть бути ефективними при наявності достатньої кількості навчальних даних для тренування моделей.

Недоліки:

– мають обмежену ефективність, коли фейкові новини імітують стиль достовірних статей;

– ігнорують соціальний контекст та поведінку користувачів, що може знижувати точність виявлення.

Методи, засновані на знаннях [38], використовують бази знань або графи знань для перевірки фактів, представлених у новинах. Автоматизовані системи

порівнюють інформацію з наявними достовірними джерелами для виявлення невідповідностей.

Переваги:

– забезпечують високу точність за наявності актуальних та повних баз знань;

– можуть виявляти фейкові новини, навіть якщо їхній стиль схожий на достовірні статті.

Недоліки:

– потребують постійного оновлення та підтримки баз знань;

– можуть бути неефективними при відсутності відповідної інформації в базах знань.

Соціально-контекстуальні методи [39] аналізують поведінку користувачів у соціальних мережах, включаючи поширення новин, взаємодії та мережеві зв'язки, для оцінки достовірності інформації. Вони враховують, як новина поширюється, хто є її основними розповсюджувачами та які реакції вона викликає. Ці методи дозволяють ідентифікувати координацію між групами акаунтів, виявляти мережі ботів або маніпулятивні кампанії, а також оцінювати загальний соціальний контекст навколо поширеної інформації.

Переваги:

– враховують динаміку поширення новин, що може підвищити точність виявлення фейків;

– дозволяють виявляти фейкові новини на ранніх етапах їхнього поширення.

Недоліки:

– потребують доступу до даних про поведінку користувачів, що може викликати питання конфіденційності та етики;

– можуть бути неефективними для новин, які ще не набули широкого поширення або мають обмежену взаємодію.

Гібридні методи [40] поєднують елементи лінгвістичних, знаннєвих та соціально-контекстуальних підходів для підвищення ефективності виявлення фейкових новин. Використовують різноманітні джерела даних та методи аналізу для комплексної оцінки достовірності інформації.

Переваги:

– забезпечують більш комплексний аналіз, що підвищує точність виявлення фейкових новин;

– можуть компенсувати недоліки окремих методів, об'єднуючи їхні сильні сторони.

Недоліки:

– складніші в реалізації та потребують більше обчислювальних ресурсів;

– можуть стикатися з проблемами інтеграції різних типів даних та методів аналізу.

Таблиця 1.1 – Порівняння підходів до виявлення фейкових новин

| **Підхід** | **Переваги** | **Недоліки** |
|---|---|---|
| Лінгвістичні | Не потребують даних про джерело; ефективні з навчальними даними | Низька точність без контексту або при імітації стилю |
| На основі знань | Висока точність за наявності баз знань | Потребують постійного оновлення баз знань |
| Соціально-контекстуальні | Враховують динаміку поширення; ефективні на ранніх етапах | Потребують даних про поведінку; етичні питання |
| Гібридні | Комплексний аналіз; поєднують сильні сторони підходів | Складні в реалізації; потребують більше ресурсів |

## 1.3 Огляд методів вирішення проблеми

У сучасному інформаційному середовищі соціальні мережі відіграють ключову роль у поширенні новин. Водночас вони стають платформою для розповсюдження неправдивої інформації, що може мати серйозні соціальні, економічні та політичні наслідки. Фейкові новини викликають дезінформацію, провокують паніку або маніпулюють громадською думкою. Для протидії цьому явищу виникає потреба у створенні ефективних програмних модулів, які здатні виявляти та класифікувати неправдиві новини.

Методи вирішення проблеми зосереджуються на трьох основних аспектах:

- аналіз контенту новин;
- виявлення патернів поширення новин у соціальних мережах;
- застосування алгоритмів машинного навчання для автоматизації виявлення фейкової інформації.

Проблема виявлення неправдивих новин у соціальних мережах є складною, оскільки потребує врахування текстових характеристик, соціального контексту поширення інформації та динаміки поведінки користувачів. Для цього використовуються різні підходи, які можна класифікувати за їхньою специфікою, метою та технологіями реалізації. Нижче розглянуто детальні методи вирішення цієї проблеми.

Лінгвістичний аналіз тексту є важливим підходом, який зосереджується на детальному вивченні стилістичних та семантичних характеристик тексту новин. Основна ідея цього методу полягає в тому, що неправдиві новини зазвичай мають характерні мовні риси, які відрізняють їх від достовірної інформації. Такі особливості можуть включати специфічні стилістичні прийоми, незвичний тон або структуру тексту, а також використання маніпулятивних або емоційно забарвлених висловів. Цей підхід дозволяє виявляти приховані патерни у мовленні, аналізувати ключові слова, які часто використовуються в неправдивих новинах, і перевіряти

текст на логічність та відповідність фактам. Наприклад, завдяки аналізу стилю написання можна ідентифікувати сенсаційні заголовки, перебільшення або маніпуляції, що часто присутні у фейкових новинах. У сукупності лінгвістичний аналіз допомагає оцінити текст на рівні його мовної структури та змісту, що робить його ефективним інструментом для ідентифікації дезінформації.

Основні методи:

– аналіз тональності тексту: виявлення емоційного забарвлення тексту (позитивний, негативний, нейтральний тон);

– семантичний аналіз: перевірка відповідності тексту загальновідомим фактам чи логічній послідовності;

– виявлення маніпулятивних висловів: визначення перебільшень, сенсаційних заголовків, закликів до емоцій.

Приклад використання:

– використання алгоритмів TF-IDF для оцінки важливості ключових слів у тексті;

– моделі на основі BERT для семантичного аналізу.

Переваги:

– простота реалізації для текстових новин.

– не потребує даних про контекст поширення новин.

Недоліки:

– складність роботи із багатомовними текстами.

– низька ефективність при обробці якісно написаних фейкових новин.

Методи перевірки фактів є ключовими у боротьбі з неправдивими новинами, оскільки вони спрямовані на автоматичну оцінку достовірності інформації, представленої в новинах. Ці методи базуються на порівнянні фактів із достовірними джерелами, такими як бази знань, структуровані дані або авторитетні медіа.

Основні методи:

– графи знань – це використання баз даних, таких як Wikidata або DBpedia, для перевірки заявлених фактів;

– автоматичне витягування фактів – це побудова запитів до баз знань для перевірки тверджень у тексті.

Приклад використання:

– використання SPARQL для витягування інформації з бази знань.

Переваги:

– висока точність за наявності актуальних баз знань;

– можливість автоматизації перевірки новин.

Недоліки:

– потребує значних обчислювальних ресурсів;

– складність підтримки актуальності баз знань.

Соціально-контекстуальний аналіз є одним із ключових підходів до виявлення неправдивої інформації, який спрямований на дослідження поведінки користувачів у соціальних мережах і аналіз способів поширення новин. Цей метод дозволяє досліджувати, як інформація циркулює між різними групами користувачів, які взаємодії відбуваються у процесі її розповсюдження, а також виявляти специфічні моделі, що можуть вказувати на дезінформацію.

Такий підхід охоплює аналіз основних вузлів у мережі, через які новини поширюються, включаючи впливових користувачів (інфлюенсерів) та автоматизовані акаунти (боти). Особлива увага приділяється динаміці поширення новин, наприклад, виявленню аномальних патернів, таких як надмірно швидке розповсюдження або незвична концентрація репостів за короткий проміжок часу. Завдяки цьому можна розкрити організовані кампанії з поширення фейкових новин або викрити джерела дезінформації.

Соціально-контекстуальний аналіз також допомагає зрозуміти, як новини впливають на громадську думку, які групи користувачів є найбільш уразливими до дезінформації, та які соціальні механізми сприяють її поширенню. Це робить цей

підхід незамінним інструментом у боротьбі з неправдивою інформацією, дозволяючи не лише виявляти фейки, а й створювати стратегії їхньої протидії.

Основні методи:

– аналіз мережевого поширення полягає у визначенні ключових вузлів, таких як інфлюенсери та боти, через які поширюються новини;

– динаміка поширення включає виявлення аномальних патернів, таких як надмірна кількість репостів за короткий час.

Приклад використання:

– графові алгоритми для аналізу мережі поширення новин;

– виявлення ботів через аналіз поведінкових патернів.

Переваги:

– виявляє ранні етапи поширення фейкових новин;

– дозволяє ідентифікувати джерела та координацію фейкових кампаній.

Недоліки:

– потребує доступу до великих обсягів даних про поведінку користувачів;

– можливі питання конфіденційності даних.

Машинне навчання є фундаментом багатьох сучасних підходів до вирішення проблеми виявлення неправдивої інформації. Ця технологія забезпечує значну гнучкість, дозволяючи ефективно працювати з різноманітними типами даних, включаючи текст, зображення та поведінкові патерни користувачів. Завдяки своїм алгоритмам і моделям машинне навчання дозволяє автоматизувати процеси аналізу, ідентифікації та класифікації даних. Це дає змогу швидко та точно виявляти ознаки фейкової інформації, навіть у складних або багатомовних контекстах.

Крім того, машинне навчання забезпечує адаптивність систем, що дозволяє враховувати нові виклики, які постійно виникають у динамічному інформаційному середовищі. Наприклад, моделі можуть бути навчені розпізнавати нові види дезінформації або аналізувати зміни в поведінці користувачів, які сприяють поширенню неправдивих новин. Завдяки цьому машинне навчання відіграє

ключову роль у створенні сучасних рішень, спрямованих на підвищення достовірності інформаційного простору.

Підходи:

– супервізоване навчання використовує позначення даних (правдиві/фейкові новини) для навчання моделей;

– глибокі нейронні мережі, такі як CNN та RNN для вивчення семантичних і стилістичних характеристик тексту;

– трансформери, зокрема моделі BERT і GPT, застосовуються для аналізу контексту і складних текстових залежностей.

Приклад використання:

– класифікація новин на основі їх текстових характеристик із використанням нейронних мереж.

Переваги:

– висока точність за наявності достатньої кількості даних;

– можливість адаптації до нових типів фейкових новин.

Недоліки:

– потреба у великих навчальних вибірках;

– високі вимоги до ресурсів для тренування моделей.

Гібридні методи поєднують переваги кількох підходів (лінгвістичного аналізу, соціального аналізу, перевірки фактів, машинного навчання) для досягнення більш комплексного й ефективного вирішення проблеми виявлення неправдивих новин. Вони інтегрують різноманітні джерела даних, моделі аналізу та алгоритми для досягнення максимальної точності та релевантності результатів. Крім того, такі методи здатні виявляти приховані патерни, які можуть залишатися непомітними при використанні окремих підходів.

Основні компоненти гібридних методів:

– лінгвістичний аналіз тексту використовується як початковий етап для попереднього виявлення підозрілих текстів, зосереджуючи увагу на таких техніках,

як токенізація, аналіз частоти слів (TF-IDF) та аналіз стилістики, що включає виявлення перебільшень або емоційного забарвлення;

– перевірка фактів інтегрується з базами знань для автоматичного порівняння тверджень із достовірними фактами, використовуючи SPARQL-запити для доступу до структурованих даних у графах знань, таких як Wikidata;

– соціально-контекстуальний аналіз передбачає дослідження патернів поширення новин у соціальних мережах, а також виявлення ботів і впливових користувачів, які можуть штучно сприяти поширенню дезінформації;

– машинне навчання передбачає використання класифікаторів для оцінки ймовірності того, що новина є неправдивою, а також застосування трансформерів, таких як BERT і GPT, для глибокого аналізу тексту та його контексту.

Приклади реалізації гібридних методів:

– комбінація BERT із графовим аналізом передбачає використання BERT для аналізу тексту новин, визначення ключових понять і тверджень, у той час як графовий аналіз оцінює, як новина поширюється мережею, та допомагає виявити аномальні патерни;

– об'єднання лінгвістичного аналізу та машинного навчання дозволяє виділити стилістичні характеристики фейкових новин за допомогою лінгвістичного аналізу, після чого модель Random Forest або нейронні мережі виконують класифікацію на основі отриманих текстових особливостей;

– інтеграція перевірки фактів і соціального аналізу передбачає зіставлення фактів із базами даних для підтвердження достовірності інформації та аналіз соціальних взаємодій з метою визначення потенційного джерела дезінформації.

Переваги:

– комплексний аналіз поєднує кілька джерел і підходів, що дозволяє охопити різні аспекти проблеми, включаючи контент, поширення та поведінку;

– адаптивність систем забезпечує можливість їх налаштування під нові патерни фейкових новин, зокрема під різні мови та формати;

– висока точність досягається завдяки використанню кількох методів, що значно зменшує ймовірність помилкової класифікації.

Недоліки:

– складність реалізації полягає в тому, що інтеграція різних методів вимагає значних зусиль у проєктуванні системи та її оптимізації;

– високі витрати пов’язані з тим, що гібридні системи потребують значних обчислювальних ресурсів і великих обсягів навчальних даних;

– проблеми масштабованості виникають через труднощі з продуктивністю при роботі з великими обсягами даних, особливо у реальному часі.

Застосування гібридних методів на практиці:

– реальні системи, такі як FactCheck.org, використовують елементи гібридних підходів для перевірки фактів і виявлення неправдивої інформації;

– наукові дослідження демонструють ефективність комбінованого аналізу тексту та соціальних взаємодій, що описується в багатьох дослідницьких статтях.

## 1.4 Постановка задачі

Для досягнення цілей наукового дослідження було встановлено мету дослідження, його об’єкт та предмет.

**Мета дослідження** – методи та засоби проектування програмних модулів для знаходження неправдивої інформації із використанням методу визначення достовірності даних.

**Об’єктом дослідження є** процеси проектування програмних застосунків для знаходження неправдивої інформації із використанням методу визначення достовірності даних.

**Предметом дослідження є** методи проектування програмних застосунків у сфері кібербезпеки та кібергігієни.

**Наукова новизна:**

– удосконалення методу проектування програмних модулів шляхом введення нових класифікаторів і проектування модулів для ефективного виявлення неправдивої інформації.

Для проведення якісного дослідження, встановлено виконання таких завдань:

– проаналізувати предметну область галузі дослідження, а саме методи, засоби та проектування модулів для знаходження неправдивої інформації;

– проаналізувати існуючі методи проектування модулів для знаходження неправдивих (фейкових новин);

– удосконалити метод проектування програмних модулів шляхом введення нових класифікаторів і проектування модулів для ефективного виявлення неправдивої інформації.

На основі запропонованого методу здійснення теоретичне проектування удосконаленого модуля, який називається FakeOut, та базується на технології BERT і досягає точності 99%. Даний модуль повинен:

– виявляти дезінформацію, зокрема, пов'язану з будь-якою неточною інформацією, а особливо такою, що виділена у наборі даних.

– бути добре оптимізовані для виявлення дезінформації, пов'язаної зокрема із військовими новинами в соціальних мережах.

– модуль FakeOut може бути розширений для загальної класифікації фейкових новин за допомогою трансферного навчання.

Практичне значення отриманих результатів. У даній кваліфікаційній магістерській роботі пропонується удосконалення методу проектування програмних модулів шляхом введення нових класифікаторів і проектування модуля для ефективного виявлення та відображення неправдивих новин потенційним використанням веб-додатку.

В подальшому результати даного дослідження можуть використовуватись при боротьбі із розповсюдженням фейкових (неправдивих) новин у різноманітних медійних засобах, зокрема у соціальних мережах.

## 1.5 Висновки до 1-го розділу

У першому розділі було проведено комплексний аналіз проблеми побудови програмних модулів для виявлення поширення неправдивих новин, розглянуто існуючі методи її вирішення та сформульовано завдання для подальшого дослідження.

Аналіз предметної області у підрозділі 1.1 показав, що поширення дезінформації у соціальних мережах є однією з найбільш актуальних проблем сучасного інформаційного суспільства. Використання соціальних мереж як основного джерела новин робить суспільство вразливим до впливу неправдивої інформації, яка здатна дестабілізувати політичні, економічні та соціальні процеси. У цьому контексті мультимодальний аналіз, штучний інтелект та нейронні мережі відкривають нові можливості для автоматизації процесів виявлення неправдивих новин. Основний акцент робиться на ранньому виявленні фейків, що дозволяє мінімізувати їхній вплив та сприяє підвищенню ефективності фактчекінгу.

Проведений у підрозділі 1.2 аналіз існуючих рішень підтвердив, що наразі розроблені системи, хоч і демонструють ефективність, все ще мають ряд недоліків. Серед них виділяються обмеженість баз знань, труднощі з інтеграцією різних методів та високі витрати на обчислювальні ресурси. Виявлено, що більшість існуючих систем використовують комбіновані підходи, які включають лінгвістичний аналіз, соціально-контекстуальний підхід та методи машинного навчання. Проте ці рішення часто потребують вдосконалення для підвищення точності та швидкості обробки даних.

У підрозділі 1.3 детально розглянуто методи вирішення проблеми. Лінгвістичний аналіз, соціально-контекстуальний підхід, перевірка фактів та машинне навчання окремо мають свої переваги, але їх поєднання у вигляді гібридних методів забезпечує найбільш комплексне та ефективне вирішення. Зокрема, використання моделей глибокого навчання, таких як BERT та GPT, у комбінації з аналізом поведінки користувачів та базами знань дозволяє досягти значних успіхів у розпізнаванні дезінформації.

У підрозділі 1.4 було сформульовано мету, об’єкт та предмет дослідження, а також окреслено ключові завдання. Основною метою є проектування модуля FakeOut, який інтегрує методи мультимодального аналізу та моделі BERT для досягнення точності 99%. Встановлено, що модуль має бути здатний до масштабування, адаптивності та ефективності у різних контекстах.

Таким чином, результати аналізу у першому розділі підтвердили доцільність і актуальність розробки програмних модулів для виявлення неправдивих новин, зокрема із застосуванням сучасних технологій обробки даних і машинного навчання. Запропонований підхід базується на інтеграції різних методів для забезпечення комплексного аналізу та підвищення якості роботи модулів. Отримані результати створюють міцну основу для подальших досліджень та розробок у рамках поставленої мети.

# 2 МЕТОД ПРОЕКТУВАННЯ ПРОГРАМНИХ МОДУЛІВ ДЛЯ ВИЗНАЧЕННЯ ДОСТОВІРНОСТІ ДАНИХ

## 2.1 Концептуальний опис наборів даних

Виявлення фейкових новин у соціальних мережах вимагає використання структурованих наборів даних, які представляють достовірну та недостовірну інформацію. Набір даних слугує основою для навчання моделей машинного навчання, що здійснюють класифікацію текстів. У даному дослідженні «справжня» інформація визначається як факт, що може бути науково підтверджений, а «фейкова» – як дезінформація, чутки або інформація, яку неможливо підтвердити. Для цього необхідно створити набори даних, які відповідають наступним вимогам:

- забезпечують якісну сегментацію текстів;
- охоплюють широкий спектр тем;
- адаптовані до сучасних викликів дезінформації.

Набори даних проходять через життєвий цикл, який включає збір, очищення, аналіз і трансформацію. Це дозволяє отримати якісну базу для роботи моделей. Зокрема, в наборі даних враховується багатоплатформенний характер джерел інформації, що охоплюють соціальні мережі, новинні портали, офіційні сайти та інші джерела.

Дослідження, проведене Pennycook та ін. [32] показує, що однією з причин поширення дезінформації є те, що люди не замислюються про правдивість контенту перед тим, як поділитися ним. Висновок, зроблений у дослідженні, полягає в тому, що якщо люди змушені думати про точність вмісту, їхній рівень «розпізнавання істини» має тенденцію до підвищення.

Створення якісного набору даних є першочерговим кроком у розробці моделі для ідентифікації фейкових новин. Набір даних слугує основою для навчання моделі, формуючи її здатність до аналізу текстів і класифікації інформації. У цьому контексті WarHub забезпечує централізоване сховище текстів, зібраних із багатьох

платформ, таких як соціальні мережі, офіційні веб-сайти, новинні портали. WarHub не лише містить актуальну інформацію, але й структурований таким чином, щоб відповідати вимогам моделей машинного навчання, забезпечуючи релевантність і різноманітність даних.

WarHub виступає як репозиторій для зберігання й аналізу текстів. Його структура забезпечує інтеграцію різноманітної інформації, включаючи:

– короткі текстові записи включають чіткі тези або заяви, що легко обробляються моделями;

– двокласова класифікація передбачає, що кожен запис отримує одну з двох міток: «Фейк» (0) або «Правда» (1), що спрощує процес аналізу та інтерпретації результатів моделі;

– тематики записів охоплюють різноманітні сфери, такі як економічні, політичні, соціальні та міжнародні питання, що забезпечує широту охоплення для ефективної ідентифікації фейкових новин у різних контекстах.

D'Ulizia та ін. [33] висуває кілька ключових характеристик наборів даних, роблячи систематичний порівняльний огляд двадцяти семи популярних наборів даних фейкових новин, серед яких домен новин, призначення програми, мова, розмір, тип новинного вмісту, рейтингова шкала та спонтанність. Таблиця 2.1 описує вищезазначені властивості WarHub.

Отримані результати, будуть запроваджуватись завдяки визначеної функції для користувачів у веб-програмі, щоб надати відгук про те, погоджуються вони чи ні з оцінкою моделі. Ця функція також забезпечить постійне і модероване розширення WarHub.

Далі виконується візуалізація даних, щоб виділити ключові характеристики WarHub. Набір даних має два стовпці – один для тексту, а інший стовпець вказує на мітку тексту. Шкала оцінок для кожного запису в наборі даних є двійковою і може бути класифікована як «Фейк» або «Справжня». У таблиці 2.1 показано приклад записів у WarHub.

Таблиця 2.1 демонструє приклад записів у WarHub.

Таблиця 2.1 – Приклад записів у WarHub

| **Шкала оцінок** | **Текст** | **Визначення неправдивості чи правдивості** |
|---|---|---|
| 0 | неготовність до вступу в Євросоюз | неправда |
| 1 | тотальна корупція в Україні | неправда |
| 2 | безрезультатність українського контрнаступу | неправда |
| 3 | Україна проти мирних переговорів, а Росія їх прагне | неправда |
| 4 | затяжна війна | неправда |
| 5 | конфлікти у вищому військово-політичному керівництві України | |

Короткі записи дозволяють уникати надмірно довгих текстів, що ускладнюють обробку, і забезпечують швидкість аналізу.

Після створення набору даних наступним кроком є формування характеристик, які допомагають моделі краще розуміти структуру, зміст і контекст тексту. Цей етап включає аналіз текстових особливостей, таких як стиль, семантика і тональність. Виділення ключових слів, визначення частот їх використання та аналіз семантичних зв’язків дозволяють моделі створювати багатовимірне уявлення про текст. Завдяки цьому підходу моделі стають здатними до глибшого аналізу складних текстів і контекстуальних зв’язків.

WarHub містить декілька основних типів даних, які дозволяють моделі ефективно працювати з різними джерелами інформації:

– категоріальні дані представляють собою тематичне розподілення записів, що охоплює такі сфери, як економічні або політичні новини. Вони допомагають класифікувати інформацію за її змістом і полегшують аналіз даних;

– текстові дані включають короткі заяви, заголовки чи тези, які легко адаптуються для перетворення у векторну форму. Це спрощує процес обробки текстів моделями машинного навчання;

– багатовимірні дані об’єднують текстову інформацію, джерела походження та класифікаційну мітку. Така структура даних забезпечує комплексний аналіз та інтеграцію різних аспектів інформації в моделі;

Ще одна ключова особливість WarHub полягає в тому, що він містить короткий текст на відміну від довгих статей. Ми виконуємо підрахунок слів для кожного запису в корпусі, підраховуючи кількість слів у стовпці «Текст». На рисунку 2.1 зображено середню довжину слова в записах у WarHub. З графіка ми бачимо, що середня довжина слова в записі менше 20 слів.

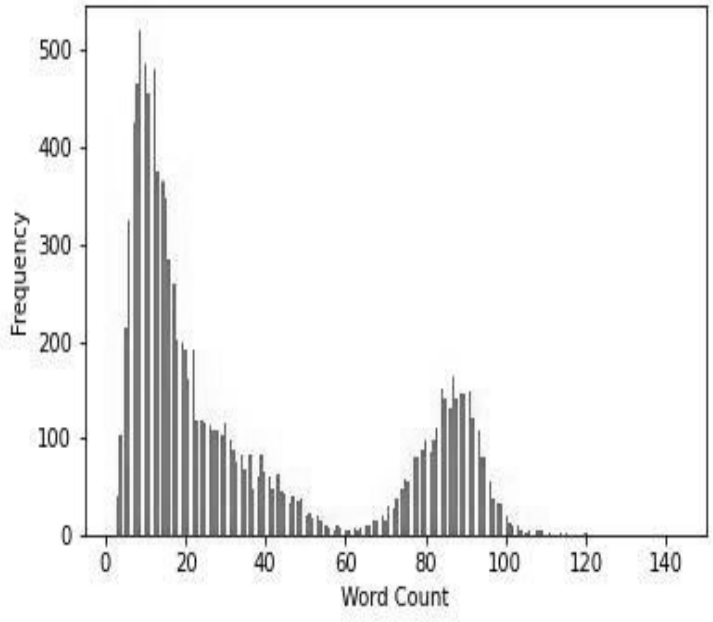



Рисунок 2.1 - Гістограма довжини речення в WarHub

Різні види фейкових новин, включаючи міфи, заголовки статей фейкових новин, неправильні переконання, хибні уявлення, містифікації та чутки. Однак ми не класифікуємо різні типи фейкових новин, оскільки WarHub охоплює

інформацію з багатьох медіаджерел. Потрібно зауважити, що неможливо визначити власний намір, коли дані збираються з різноманітних джерел.

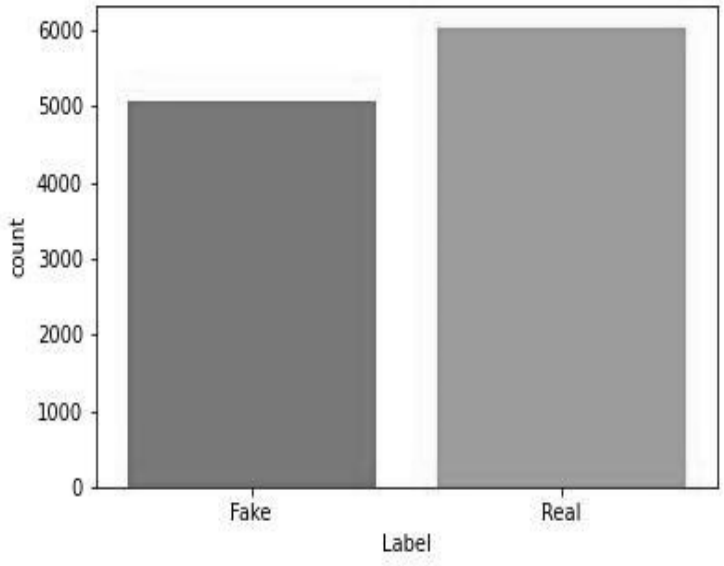



Рисунок 2.2 – Розподіл класів WarHub

Класифікатори є ядром моделі, оскільки вони виконують ключову функцію – класифікацію текстів як «Фейк» або «Правда». Для цього використовуються різні алгоритми, починаючи від традиційних, таких як Random Forest, і закінчуючи сучасними трансформерами (наприклад, BERT). Кожен класифікатор має свої переваги й обмеження, що дозволяє створити комбіновану систему для точного аналізу текстів різної складності. Поєднання кількох класифікаторів забезпечує багаторівневий підхід, який враховує як стилістичні особливості тексту, так і глибинні семантичні зв’язки.

Класифікатори є ключовим компонентом моделі. Для розділення текстів на класи використовуються такі алгоритми:

– Random Forest добре обробляє текст із середньою складністю, забезпечуючи швидку роботу з великими наборами даних. Завдяки цьому алгоритм

підходить для базового аналізу текстових характеристик, таких як стилістика і тональність;

– SVM (підтримувальні векторні машини) забезпечують високу точність класифікації, особливо в задачах із двома класами. Цей підхід ефективний для роботи з даними, які потребують точного поділу на «Фейк» і «Правда»;

– трансформери, такі як BERT і GPT, використовуються для аналізу складних текстів із багатозначним контекстом. Вони дозволяють моделі враховувати семантичні зв’язки і глиокий контекст, що забезпечує високу точність і ефективність.

Ролі класифікаторів:

– традиційні методи, як Random Forest, добре справляються з виявленням базових особливостей тексту, включаючи стилістичні характеристики і тональність;

– сучасні трансформери, такі як BERT, спеціалізуються на аналізі контексту і складних семантичних взаємозв’язків у тексті, що значно розширює можливості моделі.

## 2.2 Метод виявлення неправдивих даних

Ефективність методу виявлення неправдивих даних у соціальних мережах значною мірою залежить від якісної підготовки наборів даних, вибору алгоритмів і побудови моделі. Основною проблемою в обробці природної мови є наявність великої кількості немаркованих даних, у той час як дані з мітками — основа для навчання моделей машинного навчання під наглядом — є обмеженими. Окрім цього, аналіз існуючих наборів даних для ідентифікації фейкових новин виявляє низку проблем. Наприклад, багато наборів містять лише дані з одного джерела (твіти, статті з певного вебсайту чи соціальної мережі), що обмежує релевантність

і різноманітність інформації. Інші набори є багатомовними, що може ускладнити навчання моделей і знизити їх точність. На рисунку 2.3 демонструється відсоткове співвідношення між категоріями даних, такими як «Правда» та «Фейк».

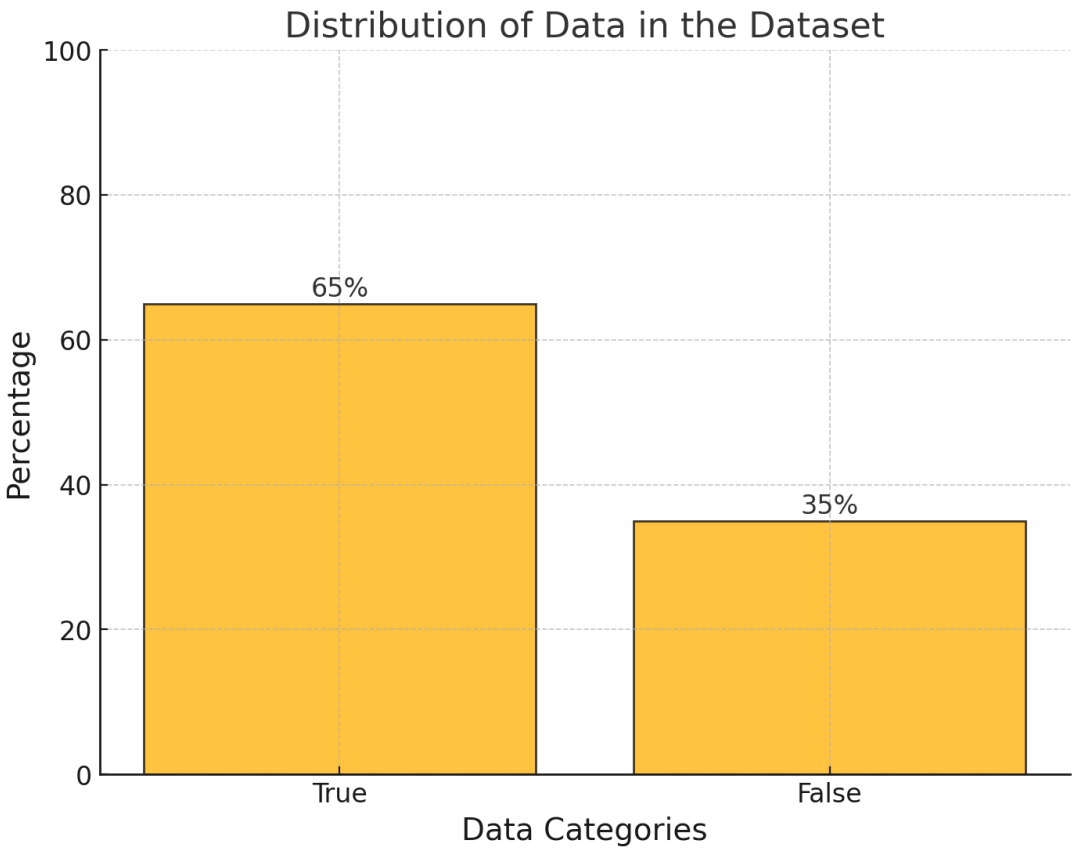



Рисунок 2.3 – Розподіл даних у наборі даних

Щоб подолати ці виклики, у даному дослідженні використовується набір даних WarHub, орієнтований на ідентифікацію фейкових новин із широким охопленням тематик і джерел. WarHub побудований для вирішення сучасних викликів у боротьбі з дезінформацією, забезпечуючи багатоплатформенність, структурованість і адаптивність до нових викликів.

Першим і ключовим етапом створення моделі для аналізу текстів у задачах обробки природної мови (NLP) є попередня обробка даних. Цей процес спрямований на усунення шумів у тексті та підготовку даних до ефективного навчання моделей машинного навчання. Ефективність класифікаційної моделі та

обґрунтованість її результатів значною мірою залежать від якості вхідних даних. Наші дослідження підтверджують, що моделі демонструють значно кращі результати при використанні попередньо обробленого тексту, ніж необробленого.

Особливо важливим є порядок виконання етапів попередньої обробки, адже він впливає на формування даних під час навчання моделі. Відсутність стандартизованих методів обробки може призвести до викривлення значення слів у тексті. Наприклад, слово «структурний» може бути скорочено до «структ», що суттєво змінює його значення та контекст у реченні. Виправлення таких недоліків є важливим завданням попередньої обробки, яке вимагає часу та ресурсів.

У процесі аналізу було виділено кілька основних етапів попередньої обробки, які забезпечують підвищення якості даних і продуктивності моделі:

– перетворення тексту в нижній регістр. Це спрощує обробку тексту, уніфікуючи великі й малі літери;

– видалення розділових знаків. Розділові знаки, як правило, не несуть змістовного навантаження для класифікаційних задач, тому їх видаляють, наприклад, за допомогою модуля string.punctuation;

– перетворення чисел на слова. Тексти можуть містити числові значення, які мають значення в контексті. Перетворення таких чисел на слова дозволяє зберегти інформацію, важливу для аналізу;

– розширення скорочень. Неформальні тексти, такі як дописи в соціальних мережах, часто містять скорочення. Їх розширення до повної форми дозволяє коректніше аналізувати текст;

– видалення посилань. Гіперпосилання у текстах можуть відволікати від основного змісту й не несуть корисної інформації для класифікації, тому їх доцільно видаляти;

– видалення зайвих пробілів і символів Unicode. Цей етап дозволяє очистити текст від технічних артефактів, які не мають семантичного значення;

– видалення стоп-слів. Стоп-слова, такі як «і», «або», «це», рідко впливають на класифікацію тексту, тому їх видаляють. Особливу увагу приділяють видаленню слів-заперечень («не», «ні»), щоб уникнути спотворення значення речень.

На рисунку 2.4, продемонстровано описаний процес, який ілюструє етапи попередньої обробки даних і взаємозв'язки між ними.

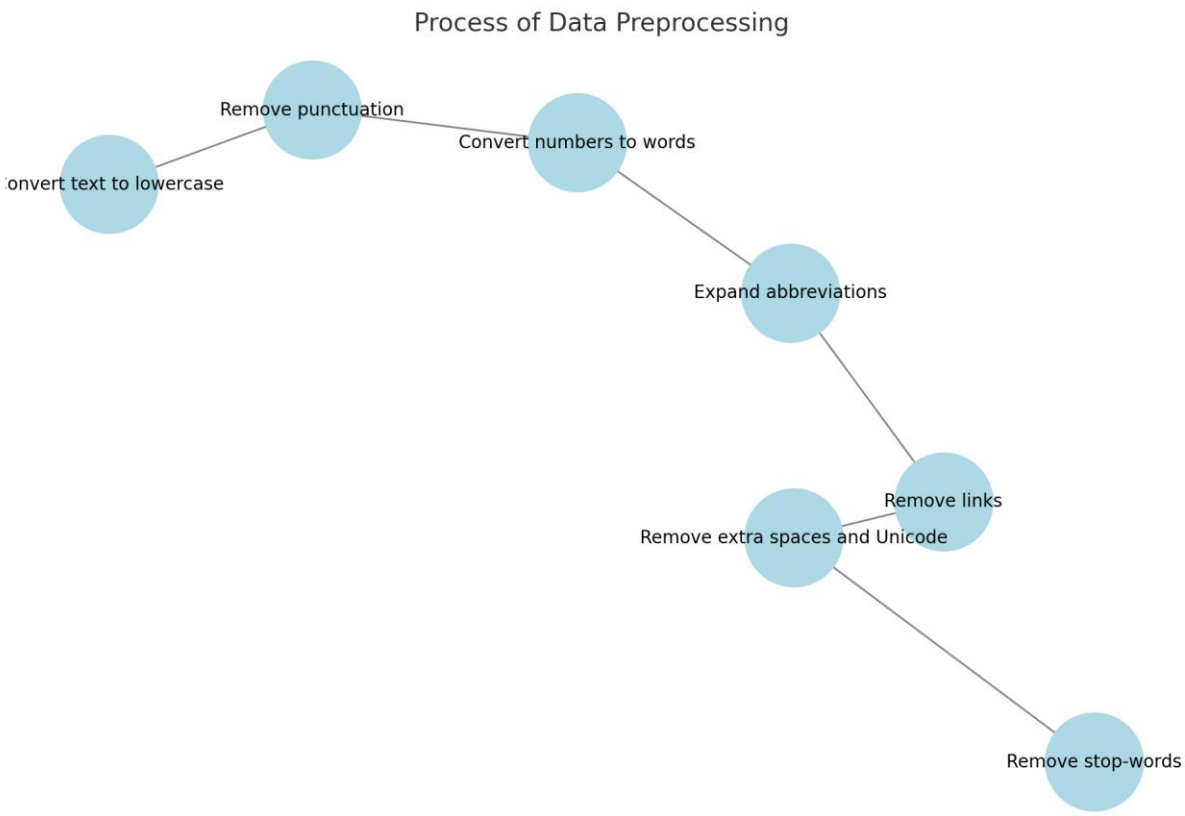



Рисунок 2.4 – Процес підготовки даних

Для навчання кожної моделі можна використовувати дві методики векторного представлення – TF-IDF (термін частотно-інверсної частоти документа) і Count Vectorizer.

Частота термінів – зворотна частота документів (TF-IDF) – це широко використовуваний статистичний метод в обробці природної мови та пошуку інформації. Він вимірює, наскільки важливим є термін у документі по відношенню до колекції документів (тобто по відношенню до корпусу).

Слова в текстовому документі перетворюються на числа важливості за допомогою процесу векторизації тексту. Існує багато різних схем оцінювання векторизації тексту, однією з найпоширеніших є TF-IDF.

Як випливає з назви, TF-IDF векторизує (оцінює) слово шляхом множення частоти терміна (TF) на зворотну частоту документа (IDF).

Частота терміна: TF терміна або слова - це кількість разів, коли термін з'являється в документі, порівняно із загальною кількістю слів у документі.

Зворотна частота документа: IDF терміна відображає частку документів у корпусі, які містять цей термін. Слова, унікальні для невеликого відсотка документів (наприклад, терміни технічного жаргону), отримують вищі значення важливості, ніж слова, поширені в усіх документах.

У перекладі на українську мову важливість терміну висока, якщо він часто зустрічається в одному документі і рідко - в інших. Коротше кажучи, поширеність терміна в документі, виміряна за допомогою TF, врівноважується рідкістю між документами, виміряною за допомогою IDF.

TF-IDF – це проста, але потужна техніка, заснована на підході Bag of words для векторизації тексту. Це статистичний спосіб обчислення релевантності терміна в документі. Він працює шляхом обчислення добутку ваги терміна в документі, який називається частотою терміну (TF) (підраховує кількість разів, коли слово з’являється в документі, поділену на загальну кількість слів у документі) та зворотну частоту документа (IDF). IDF обчислюється шляхом поділення загальної кількості записів або речень у корпусі на кількість записів або речень, що містять слово. IDF збільшує вагу слів, які рідко зустрічаються в корпусі документа, замість термінів, які часто зустрічаються. TF-IDF перетворює необроблений текст у розріджену двовимірну матрицю характеристик TF-IDF розміру, кількості записів у корпусі, загального словникового запасу корпусу. Обмеження використання TF-IDF полягає в тому, що результуюча матриця не може захопити жодної семантичної інформації. Цей метод може бути обчислювально дорогим у випадку великого словникового запасу, оскільки розмір матриці залежить від кількості унікальних

слів у корпусі. З математичної точки зору, TF-IDF для слова w у реченні s із повного набору записів ( WarHub ) D обчислюється як:

$$TF - IDF\ (w, s, D) = TF\ (w, s) * IDF\ (w, D) \quad (1)$$

де TF (w, s) = кількість разів, коли слово w з'являється в реченні s / загальна кількість слів у реченні s, а IDF (w, D) = log ( кількість речень s у корпусі D / кількість речень s у корпусі D, що містить слово w)

Графічна векторизація – ще один простий прийом частотного представлення слів. Це дуже схоже на одноразове кодування. Ця техніка також створює розріджені матриці, такі як векторизація TF-IDF. Розмір матриці залежить від кількості записів у корпусі та загального словникового запасу. Цей простий прийом також не в змозі охопити реляційну інформацію слів у реченні.

Класифікація тексту є основою методу виявлення неправдивих даних, і для її реалізації використовуються різні алгоритми машинного навчання, кожен із яких має свої особливості, переваги та обмеження. Нижче розглянуто основні алгоритми, які застосовуються в дослідженні, з описом їхніх принципів роботи та переваг.

KNN – це простий непараметричний алгоритм класифікації тексту. Він працює, знаходячи найближчого сусіда вектора та обчислюючи подібність між найближчими векторами. На рішення щодо класифікації нового вектора впливає значення K і метрика відстані. У літературі є кілька спроб використання KNN для класифікації тексту. Основним недоліком KNN для класифікації тексту є те, що він чутливий до розподілу даних і висока часова складність. Щоб ідентифікувати найближчих сусідів, він має обчислити відстань вектора до всіх існуючих векторів у наборі даних. KNN працює добре, якщо є невелика кількість записів. Можна використовувати бібліотеки, наприклад scikit для впровадження KNN. Також мона експериментувати з різними значеннями K за допомогою Grid search CV і виявити, що надзвичайно мале значення K, наприклад, K менше десяти, призводить до

перезапуску. Підтримуючи значення K між 10 і 100, досягаються майже подібні результати, але зі збільшенням значення K збільшується й часова складність. Збільшення значення K понад 100 призводить до поганих результатів. Ми вибрали значення K рівним 10. Наша модель KNN добре працює з TF-IDF порівняно з Count Vectorizer і забезпечує точність навчання та тестування 85,5% і 79,4% відповідно.

Логістична регресія – це дискримінаційний класифікатор і добре працює з бінарною класифікацією. Він заснований на концепції ймовірності та використовує сигмоподібну функцію для обмеження виходу між 0 і 1. Встановлюється поріг для прогнозування фактичного класу; потім оцінена ймовірність перетворюється на клас залежно від порогового значення. Також вчені експериментували з логістичною регресією разом із різними моделями, такими як Дерево рішень, SVM, Випадковий ліс і глибокі моделі для ідентифікації фейкових новин щодо COVID-19. Автори експериментують з іншими методами векторизації TF-IDF, і їх найкраща модель логістичної регресії отримана за допомогою триграми з точністю 81%. В даному дослідженні розглядається модель логістичної регресії за допомогою TF-IDF і Count Vectorizer. Обидві моделі однаково добре себе показують на тестовому наборі. Однак модель, навчена за допомогою TF-IDF, має тенденцію змінювати дані. Збільшення значення C також призводить до зміни, і, отже, необхідно зберенти значення C за замовчуванням рівним 1. Після експериментів з різними значеннями регулярізації найкращі результати отримані з регуляризацією L2. Наша найефективніша модель логістичної регресії забезпечує точність навчання та тестування 92,85% і 86,4% відповідно.

Наївний Байєс – це імовірнісний і генеративний класифікатор. Він заснований на теоремі Байє . Naive Bayes не враховує порядок слів і, отже, не зберігає жодного зв’язку між словами під час класифікації. Це один із перших алгоритмів, який вирішив проблему фільтрації спаму. Різноманітні версії наївного байєса, такі як Gaussian NB, Multinomial NB і Bernoulli NB, доступні та використовувалися для класифікації тексту в минулому. Гранік і Месюра [13] пропонують простий підхід для класифікації фейкових новин за допомогою

алгоритму Naive Bayes. Дослідження використовує 2000 записів зі стандартного набору новин BuzzFeed, і їх модель досягає точності тесту в 74%. Можна використовувати Multinomial NB для експерименту, оскільки він підходить для частоти таких функцій, як кількість слів.

SVM – це універсальний двійковий класифікатор, який працює, знаходячи лінійну або поліноміальну порогову функцію, щоб відокремити екземпляри одного класу від решти. Він добре підходить для обробки великого простору функцій, наприклад тексту, і добре працює при класифікації тексту порівняно з іншими алгоритмами. У літературі є докази того, що моделі SVM перевершують більшість інших моделей для завдань класифікації тексту. Наприклад, Abdelminaam et al. [13] показують, що модель SVM, навчена за допомогою уніграм TF-IDF, найкраще розпізнає фальшиві та справжні твіти, що містять інформацію про COVID-19, порівняно з іншими простими моделями, такими як Дерево рішень, KNN, Випадковий ліс і Логістична регресія. Модель SVM досягає точності тесту 96,38%. Ми навчаємо модель SVM за допомогою бібліотеки sklearn . SVM надає вибір таких ядер, як лінійне, поліноміальне, гауссове, RBF і сигмоподібне. Лінійне ядро часто є кращим вибором порівняно з іншими нелінійними ядрами через менші параметри для гіперналаштування. Ми намагаємося знайти найкращу комбінацію гіперпараметрів C, гами та ядра за допомогою GridSerachCV із 5-кратною перехресною перевіркою. Найкраща точність досягається за допомогою ядра RBF і значень гамми та C, встановлених на 0,1 і десять відповідно. Однак ця комбінація призводить до переобладнання, і, отже, ми зменшуємо значення гами до 0,01. Ефект переобладнання був більшим у моделі векторизатора Count, ніж у векторизаторі TF-IDF. Отже, ми зменшуємо значення C до п’яти та додатково зменшуємо значення гами до 0,007 у SVM за допомогою векторизатора Count. Зменшення значення гами вирішує проблему переобладнання, але дещо знижує точність. Низьке значення гамми та C призводить до недообладнання, а високе значення призводить до переобладнання. Остаточна точність тренування та

тестування для TF-IDF і Count Vectorizer SVM становить 89%, 85% і 91%, 83% відповідно.

Класифікатор MLP – це базова нейронна мережа прямого зв'язку, яка має конфігурацію вхідного рівня, за яким слідує прихований рівень, за яким слідує вихідний рівень. Ми використовуємо просту реалізацію MLP, яка має багато параметрів, як-от кількість прихованих шарів, функцію активації, значення альфа для регулярізації L2, ітерації, що позначають епохи, і алгоритм оптимізації, і швидкість навчання. Ми експериментуємо з різними значеннями параметрів і спостерігаємо, що гіпернастроювання параметрів може досягти майже 99% точності навчання. Однак моделі, як правило, переповнюють і мають низькі показники на тестовому наборі даних. Щоб вирішити проблему надмірного оснащення, необхідно збільшити значення альфа до 1,5. Альфа - термін регуляризації L2; збільшення його значення заохочує малі ваги, що призводить до межі рішення, яка менше відповідає даним. Ми також надаємо можливість ранньої зупинки. Увімкнувши це, алгоритм залишає 10% даних для перевірки та зупиняє процес навчання, якщо не бачить подальшого покращення оцінки перевірки. Ми зберігаємо швидкість навчання адаптивною, а функцію активації за замовчуванням (ReLU). З цими значеннями параметрів і одним прихованим шаром зі 100 нейронами модель досягає 87% і 83% точності навчання і тестування з векторизацією TFIDF і 92%, 85% точності навчання і тестування з векторизацією Count.

Модель на основі BERT. У 2018 році Google опублікував свою попередньо підготовлену модель BERT (Bidirectional Encoder Representations from Transformers). BERT є попередньо підготовленою моделлю, тому її словниковий запас є фіксованим. Його навчають на основі статей і книжкового корпусу Вікіпедії, а його розмір слів становить близько 30 000. Нам потрібно використовувати токенізатор BERT, щоб зіставити слова з послідовністю вбудовування для використання BERT. BERT розгортає WordPiece Tokenizer, алгоритм токенізації на основі підслів. Якщо слово не існує в словнику BERT, воно

розбивається на підслова та утворює лексеми. Кожне слово зіставляється з ідентифікатором словника. Вбудовування слів мають контекстний характер і є такими, що відстань між векторами відображає їхню подібність. Вхідні дані для моделі BERT мають фіксований формат.

Усі речення, що передаються до BERT, мають бути фіксованої довжини, доповнені або скорочені. Вибір максимальної довжини вхідних даних впливає на час навчання та точність. Вхідне речення також має починатися та закінчуватися штучними лексемами «CLS » та «SEP» відповідно. Лексема CLS позначає початок речення. Другим вхідним елементом моделі BERT після вхідних ідентифікаторів (слів, відображених у токени) є маска уваги. Це масив 0 та 1 розміру фіксованої довжини, що розрізняє фактичні слова та доповнені маркери. Це допомагає моделям зрозуміти, яким словам слід віддавати пріоритет, а які можна ігнорувати. BERT доступний у двох варіантах – базовому та великому. Різниця полягає в кількості трансформаторних шарів, головок уваги та прихованих блоків.

Результати токенізера BERT, застосовані до одного зразка запису. Перший рядок на малюнку показує текст. На наступному кроці слова в реченні розбиваються на лексеми та додаються з CLS і SEP на початку та в кінці. Результатом роботи токенізера є словник із масивами вибраного розміру фіксованої довжини. Перший масив – це вхідні ідентифікатори, які є масивом токенів, зіставлених із числами. Масив починається і закінчується номерами 101 і 102, що представляють CLS і маркер SEP. Інший масив – це маска уваги з 0 і 1, де 0 представляє маркери заповнення, а 1 позначає справжні маркери.

Можна використовувати існуючу модель BERT у багатьох завданнях НЛП, таких як класифікація тексту, розпізнавання імен і сутностей, позначення частин мови та відповіді на запитання, і скористатися перевагами попереднього навчання, а не навчання моделі з нуля. У літературі досліджено декілька типів моделей класифікації тексту на основі BERT. Ding та ін. [21] представляють розумову модель класифікації фейкових новин на основі BERT. Ця модель забезпечує кращі результати, ніж сучасні моделі для набору даних LIAR. Чоудхарі та ін. [23]

запропонував «BerConvNet», гібридну модель глибокого навчання, що поєднує BERT і CNN. Дослідження підтверджує ефективність BerConvNet на чотирьох різних контрольних наборах даних для ідентифікації фейкових новин. Порівняльний аналіз, проведений у дослідженні, показує, що нова модель перевершує існуючі найсучасніші моделі за різними показниками ефективності.

## 2.3 Алгоритм процесу визначення неправдивої інформації

Поширення неправдивої інформації є складним процесом, що підживлюється використанням нечесних суб'єктів, які діють незалежно або на основі широкомасштабне використання мережі ботів соціальних мереж. Обидва обманюють читачів, створюючи ілюзію консенсусу щодо неправдивої інформації, наприклад, повторивши її кілька разів або висловивши пряму підтримку для цього. Ці облікові записи мають на меті штучно створити вірусність свого вмісту (наприклад, шляхом «проголосування»/просування вміст на ранній стадії), щоб поширювати повідомлення з неправдивою інформацією ще швидше та глибше, ніж правдива інформації.

Самотні вовки працюють, створюючи кілька підроблених облікових записів «sockpuppet» або «sybil» і використовуючи їх у координація, щоб відобразити ту саму точку зору, шляхом написання подібних відгуків на платформах електронної комерції або створення подібні коментарі на публічних форумах. Операції самотнього вовка з використанням кількох облікових записів можуть бути особливо переконливими оскільки читачі, як правило, не знають, що вся дискусія є сфабрикованою та фактично походить від одного джерело.

Наприклад, в онлайн-розмовах Kumar et al. [17] характеризують цю поведінку, вивчивши 61 мільйон коментарі, зроблені 2,1 мільйонами користувачів на кількох дискусійних платформах. Вони виявили, що в той час як sockpuppets

можуть використовуватися з доброзичливими намірами, sockpuppets з оманливими намірами зустрічаються вдвічі частіше. Оманлива sockpuppets відповідають один одному згодою та підтримкою, і негативно ставляться до облікових записів, які не згодні. Крім того, ці облікові записи займають центральне розташування в мережі зв'язку, а отже, у ключових точках поширювати неправдивий вміст.

Так само облікові записи sybil в комунікаціях і соціальних мережах створюються для інтеграції добре проникають у мережу та запобігають виявленню, щоб збільшити свій вплив на інших [13]. У більшому масштабі ботнети соціальних мереж використовуються для поширення неправдивої інформації. Боти, які є підробленими або скомпрометованими облікові записи, контрольовані окремою особою або програмою, використовуються для двох основних цілей: надсилання того самого швидко донести інформацію до широкої аудиторії та завищити «соціальний статус» певних користувачів, обидва з яких роблять неправдивими щоб інформація виглядала надійною та законною.

Бессі та ін. [13] і Шао та ін. [17] вивчали використання ботів у політичних кампаніях і знайшли що облікові записи ботів відповідальні за майже одну п'яту всіх політичних балачок у Twitter, і що неправдива інформація більша ймовірність поширюватися ботами, ніж реальними користувачами. 25% неправдивої інформації твітів створено ботами.

Загальна стратегія, яку використовують боти, полягає в тому, щоб націлювати інформацію на більше впливові реальні користувачі, які іноді можуть піддаватися впливу та повторно ділитися неправдивим повідомленням, пересилаючи його ширшому колі аудиторія. Щоб підвищити «соціальний статус», оператори ботнетів пропонують послуги, які надають фальшивих підписників використовуючи своїх ботів, щоб стежити за платними обліковими записами клієнтів.

Шах та ін. [15] досліджували ці служби та виявили, що вони працюють за моделями «freemium» і «premium», де перша складається з скомпрометованих або справжніх облікових записів користувачів і останній складається з підроблених

облікових записів або акаунтів ботів. Ці дві моделі діють досить чітко - шахрайство з безкоштовним доступом облікові записи створюють кліки з високою щільністю підписаних облікових записів, які торгують підписниками між собою, тоді як преміум шахрайські облікові записи створюють щільні двосторонні ядра, тобто один набір облікових записів слідує за клієнтами, що платять.

Це збільшується очевидна надійність користувачів, які потім можуть бути використані для подальшого поширення неправдивої інформації. У недавньому дослідженні Vosoughi та ін. [15] проаналізували понад 126 000 каскадів неправдивої інформації в Twitter протягом період 11 років і показав, що люди відповідальні за поширення неправдивої інформації в Twitter, а не ботів.

Використовуючи інструмент виявлення ботів Twitter BotOrNot, розроблений Davis et al [21], вони ідентифікували бота та облікові записи неботів, які займалися поширенням неправдивої інформації. Вони виявили, що в Твіттері були люди, а не боти відповідальність за поширення неправдивої інформації, оскільки боти відповідали за прискорення поширення як правдивої інформації і неправдивої інформації приблизно порівну.

Навіть після видалення активності бота спостерігалося поширення неправдивої інформації далі, глибше, швидше та ширше, ніж справжня інформація. Крім того, вони виявили, що не-бот облікові записи на Твіттер, відповідальний за поширення неправдивої інформації, був новішим, мав менше підписників і фоловерів були менш активними.

Хоча це стосується Twitter, інші платформи можуть поводитися по-іншому, а також поширення нечесних осіб у створенні та поширенні неправдивої інформації є поширеним явищем.

Таким чином, використання маріонеток і ботнетів використовується для розповсюдження неправдивої інформації у великій кількості реальних користувачів у соціальних мережах. Ці облікові записи працюють за допомогою підроблених і комп’ютеризованих облікових записів для збільшення видимість неправдивої інформації та соціальний статус акаунтів, які її поширюють.

Алгоритм визначення неправдивої інформації ґрунтується на поетапному та багаторівневому аналізі даних, які отримуються із соціальних мереж. Цей процес охоплює різні аспекти взаємодії з інформацією та обліковими записами, зосереджуючи увагу на ідентифікації джерел дезінформації та їхніх способів поширення через структурований підхід, що включає кілька послідовних етапів, кожен із яких виконує свою унікальну функцію.

– збір даних із соціальних мереж включає текстові повідомлення, метаінформацію акаунтів (кількість підписників, активність) та історію взаємодії;

– виявлення активних облікових записів здійснюється шляхом ідентифікації акаунтів, які можуть бути ботами, sockpuppet- або sybil-акаунтами, через аналіз їхньої поведінки, зокрема частоти публікацій чи нетипових взаємодій;

– аналіз шаблонів поведінки передбачає виявлення повторення повідомлень, шаблонів вірусного поширення та координації між обліковими записами з метою ідентифікації ознак організованої діяльності;

– виявлення координованих дій включає оцінку того, чи група облікових записів діє узгоджено для створення ілюзії підтримки певної інформації, наприклад, шляхом поширення ідентичних повідомлень або взаємодії з однаковим контентом;

– перевірка вмісту здійснюється шляхом аналізу тексту на відповідність фактологічним базам даних, стилістичним характеристикам написання, використанню емоційних тригерів та наявності маніпулятивної мови;

– оцінка достовірності інформації проводиться на основі зібраних даних, враховуючи її відповідність правдивим фактам та наявність зв’язку з підозрілими обліковими записами;

– інформація, яка має низький рівень достовірності, позначається як потенційно неправдива для подальшої перевірки;

– система здійснює оповіщення модераторів платформи для перевірки контенту або автоматично вживає заходів, таких як обмеження поширення повідомлення.

Цей алгоритм розроблений для комплексного та систематичного виявлення неправдивої інформації, враховуючи складність і багатогранність сучасних викликів у цифровому середовищі. Він спрямований на протидію різним формам дезінформації, які активно поширюються за допомогою сучасних технологій, таких як боти, sockpuppet-акаунти та автоматизовані ботнети

На рисунку 2.5 продемонстровано алгоритм визначення неправдивої інформації.

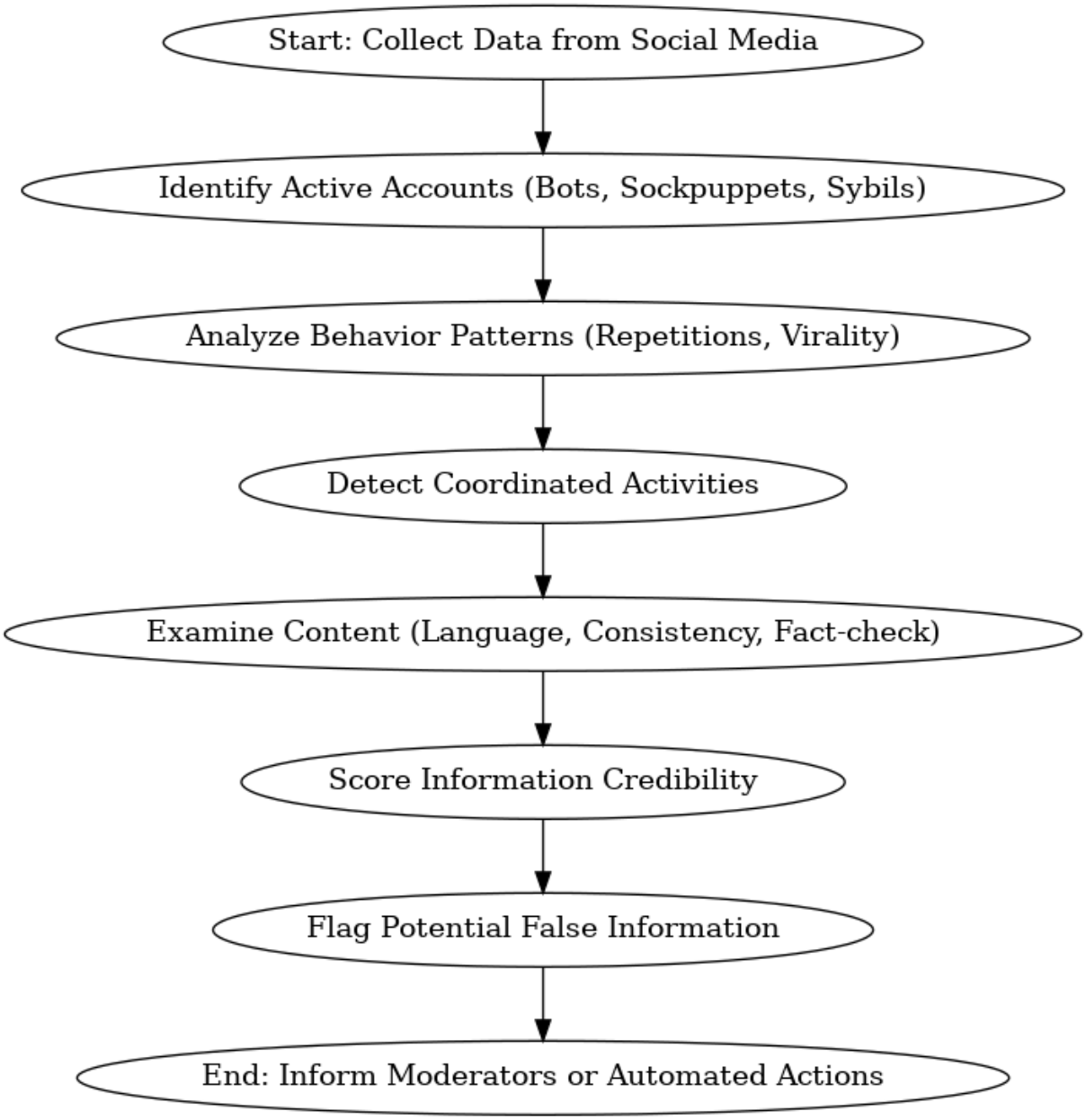



Рисунок 2.5 – Алгоритм визначення неправдивої інформації

## 2.4 Висновки до другого розділу

Отже, у другому розділі було встановлено, що розробка та впровадження модулів для виявлення неправдивої інформації в соціальних мережах є важливим кроком у боротьбі з дезінформацією. Ключову роль у цьому процесі відіграють якісно підготовлені набори даних, які є основою для навчання моделей машинного навчання. Набори даних повинні відповідати високим стандартам, охоплюючи широкий спектр тем, джерел та форм представлення інформації. Інструмент WarHub демонструє потенціал у створенні структурованої платформи для збору, аналізу та класифікації інформації, враховуючи багатоплатформенний характер джерел та адаптивність до сучасних викликів.

Ефективність моделей залежить не лише від даних, але й від правильної архітектури алгоритмів класифікації. Поєднання традиційних підходів, таких як Random Forest та SVM, із сучасними трансформерами на базі BERT дозволяє створити потужні рішення для аналізу складних текстів. Попередня обробка даних, включаючи очищення тексту, видалення шумів та нормалізацію, є важливим етапом для забезпечення достовірності результатів.

Алгоритм визначення неправдивої інформації враховує як поведінкові шаблони користувачів, так і вплив штучних елементів, таких як боти та маріонеткові облікові записи. Багаторівневий підхід дозволяє аналізувати як вміст тексту, так і способи його поширення, зокрема через ботнети чи мережі «sockpuppet»-акаунтів. Це сприяє створенню інтегрованої системи, здатної ефективно ідентифікувати неправдиву інформацію.

Запропоновані рішення демонструють, що модульний підхід, поєднаний із багаторівневим аналізом, дозволяє значно підвищити точність виявлення фейкових новин. Це відкриває перспективи для забезпечення інформаційної безпеки, підвищення довіри до інформаційного простору та підтримки громадської свідомості в умовах цифрового середовища.

# 3 АРХІТЕКТУРА СИСТЕМИ ВИЯВЛЕННЯ НЕПРАВДИВИХ ДАНИХ

## 3.1 Формування та аналіз вимог програмної реалізації веб-системи визначення достовірності даних

Визначення вимог до програмної реалізації є одним із найважливіших етапів у процесі проектування та розробки будь-якого програмного забезпечення. Цей процес формує основу для створення продукту, що відповідає очікуванням користувачів і задовольняє бізнес-цілі. На цьому етапі закладаються фундаментальні характеристики майбутньої системи, які будуть реалізовані в процесі розробки.

Одним із ключових завдань у цьому процесі є визначення обсягу продукту, тобто які функціональні можливості та характеристики мають бути реалізовані. У цьому контексті процес визначення вимог стає центральною діяльністю, яка включає збір, аналіз, опис і встановлення пріоритетів вимог. Такий підхід дозволяє забезпечити чітке розуміння очікувань зацікавлених сторін і створити план реалізації, що враховує всі необхідні аспекти.

Вимоги до програмного забезпечення зазвичай поділяються на дві основні категорії: функціональні та нефункціональні.

Функціональні вимоги описують конкретні послуги або функції, які має надавати програмне забезпечення. Вони визначають:

- які операції має виконувати система;
- які відповіді повинні генеруватися на основі певних вхідних даних;
- як система повинна поводитися в різних сценаріях використання.

Наприклад, для системи виявлення неправдивих новин функціональні вимоги можуть включати:

- здатність аналізувати текстові дані для визначення ймовірності фейковості;

– інтеграцію з API соціальних мереж для збору реальних даних у реальному часі;

– створення звітів із результатами класифікації.

Нефункціональні вимоги є важливим компонентом процесу визначення вимог до програмного забезпечення, адже вони описують загальні характеристики системи, що впливають на її якість і ефективність, але не пов’язані безпосередньо з конкретними функціями чи модулями. Вони забезпечують належне функціонування системи в реальних умовах, враховуючи технічні, експлуатаційні та користувацькі аспекти.

Продуктивність системи визначає її здатність швидко та ефективно виконувати завдання в умовах різних навантажень. Це поняття охоплює такі аспекти, як час відповіді на запити користувача, тривалість обробки великих обсягів даних і виконання складних операцій. Забезпечення високої продуктивності вимагає оптимізації програмного коду, використання ефективних алгоритмів, налаштування апаратного забезпечення та вибору відповідної архітектури. Наприклад, у системах бізнес-аналітики критично важливо обробляти великі набори даних у реальному часі, щоб забезпечити оперативність прийняття управлінських рішень. Низька продуктивність може призводити до зниження довіри користувачів та втрати конкурентоспроможності.

Масштабованість визначає здатність системи адаптуватися до зростання навантаження або зміни потреб користувачів без втрати її продуктивності. Це може бути реалізовано за рахунок масштабування "вгору" (додавання потужніших апаратних ресурсів) або "вшир" (розширення кількості серверів чи вузлів у кластері). Наприклад, якщо система обробляє потік інформації, що постійно зростає, важливо забезпечити безперебійну роботу шляхом додавання нових серверів або використання хмарних рішень. Масштабованість також включає можливість інтеграції нових функціональних модулів, розширення баз даних чи забезпечення сумісності з іншими системами.

Безпека передбачає захист даних користувачів від несанкціонованого доступу, витоків чи змін. Це досягається завдяки впровадженню механізмів шифрування даних, багаторівневої аутентифікації, регулярного аудиту безпеки та дотримання стандартів, таких як GDPR чи ISO/IEC 27001. Особлива увага приділяється системам, які працюють із конфіденційною або персональною інформацією.

Юзабіліті спрямоване на забезпечення інтуїтивно зрозумілого та зручного інтерфейсу для кінцевих користувачів. Основними аспектами є простота навігації, зрозумілі повідомлення про помилки та адаптація інтерфейсу до потреб користувачів з обмеженими можливостями. Мета полягає у створенні комфортного та доступного програмного забезпечення для всіх категорій користувачів.

Сумісність стосується здатності системи інтегруватися з іншими платформами, додатками чи інфраструктурами. Це включає підтримку різних операційних систем, сумісність із популярними базами даних або API та відповідність стандартам галузі. Забезпечення високої сумісності сприяє легкому впровадженню системи в існуюче технічне середовище.

Загальне значення нефункціональних вимог. Ці вимоги забезпечують надійність, масштабованість, безпеку, зручність та інтеграцію системи, що гарантує її відповідність сучасним викликам і потребам користувачів. Вони також слугують критеріями оцінювання якості розробленого програмного забезпечення та його здатності ефективно працювати в різних умовах.

Узгодження вимог між усіма зацікавленими сторонами, включаючи розробників, бізнес-аналітиків, кінцевих користувачів та менеджерів, є важливим етапом у процесі розробки програмного забезпечення. Такий підхід сприяє уникненню конфліктів, які можуть виникнути через різне бачення продукту, та забезпечує, щоб кінцевий продукт максимально відповідав очікуванням усіх учасників процесу.

Процес визначення вимог має бути достатньо гнучким, щоб враховувати можливі зміни в умовах бізнесу чи технологічного середовища. У таких випадках

ітеративний підхід стає надзвичайно корисним, адже він дозволяє поступово уточнювати вимоги та адаптувати їх на основі отриманих результатів під час проміжних етапів розробки. Це дає змогу швидко реагувати на нові виклики чи уточнення, зберігаючи актуальність і релевантність вимог.

Чітко сформульовані та структуровані вимоги є фундаментом для створення ефективного, стабільного та відповідного до потреб користувачів програмного забезпечення. Вони забезпечують єдність у розумінні цілей проекту, спрощують процеси розробки й тестування та сприяють отриманню кінцевого продукту, який відповідає запитам ринку й очікуванням клієнтів.

У Таблиці 3.1, 3.2 показано функціональні вимоги до двох прикладних модулів. Кожна вимога представлена кодом, описом і її пріоритетом (суттєвим, важливим або бажаним) для кінцевого продукту. З іншого боку, таблиця 3.3 показує нефункціональні вимоги.

Таблиця 3.1 – Функціональні вимоги для проектованого модуля аналізу інформації

| **Опис** | **Пріоритет** |
|---|---|
| Система повинна мати можливість аналізувати посилання на новини, які містять таку інформацію, як: назва, автор і дата публікації. | Необхідний |
| Система може аналізувати інформацію різними мовами | Необхідний |
| Кожна представлена інформація повинна супроводжуватися поясненням того, як вона може допомогти вашому аналізу. | Необхідний |
| Пояснення має бути чітким і зрозумілим для всіх, навіть для тих, хто не має глибоких знань про віртуальне середовище. | Важливий |
| Наступні критерії повинні бути показані для певної новини, розділені відповідно до ключових питань, які допомагають користувачеві прийняти рішення: | Необхідний |

Продовження Таблиці 3.1

| | |
|---|---|
| Кожна представлена інформація повинна супроводжуватися поясненням того, як вона може допомогти вашому аналізу. | Необхідний |
| Пояснення має бути чітким і зрозумілим для всіх, навіть для тих, хто не має глибоких знань про віртуальне середовище. | Важливий |
| Наступні критерії повинні бути показані для певної новини, розділені відповідно до ключових питань, які допомагають користувачеві прийняти рішення: | Необхідний |
| • Хто стоїть за новиною?<br>– відповідальний за домен запису;<br>– автор;<br>– інформація про транспортний засіб отримана від Вікіпедія.<br>• Які докази для позову?<br>– публікації, пов’язані з історією перевірки фактів агентства;<br>– дата публікації.<br>• Що кажуть інші джерела?<br>– подібні історії в інших джерелах новин. | Необхідний |
| Якщо будь-яка інформація з певних причин недоступна/неможливо отримати доступ, слід представити користувача із заохочувальним повідомленням, пояснюючи, що отримати інформацію неможливо, але вони можуть знайти її за допомогою інших засобів, таких як пошукові системи або агрегатори новин, наприклад. | Важливий |

Таблиця 3.2 – Функціональні вимоги для проектованого модуля аналізу інформації

| **Опис** | **Пріоритет** |
|---|---|
| Користувачеві має бути представлено декілька новин | Необхідний |
| Новини мають бути представлені у макетах, які імітують віртуальні контексти, такі як *дописи* з соціальних мереж, новинних сайтів або повідомлення з комунікаційних програм. | Важливий |
| Представлена новина повинна містити, окрім тексту, таку інформацію, як: назва засобу публікації, дата публікації, ім’я автора, *посилання* , за яким новина була опублікована. | Необхідний |
| Якщо аналіз правильний, система має надати користувачеві відгук. | Необхідний |
| У разі неправильного аналізу система повинна бути в змозі представити користувачеві інформацію, яку слід було взяти до уваги для правильного аналізу. | Необхідний |
| Наприкінці користувачеві має відобразитися повідомлення зворотного зв’язку. Якщо ви досягли результату, меншого або рівного 50%, вам має бути показано підбадьорливе повідомлення, а також поради щодо того, що робити, щоб реалізувати свої навички. У разі досягнення результату, що перевищує 50%, має відобразитися вітальне повідомлення. | Необхідний |

Таблиця 3.3 – Нефункціональні вимоги для проектованого модуля аналізу інформації

| **Опис** | **Пріоритет** |
|---|---|
| Система має бути сумісною з сучасними браузерами та не мати відмінностей у поведінці між цими браузерами. | Необхідний |
| Інструмент також має бути сумісний із мобільними браузерами. Таким чином, оперативність є аспектом, який слід враховувати в процесі впровадження. | Важливий |
| Критерії для аналізу повинні бути викладені таким чином, щоб їх було легко читати користувачеві, і, крім того, вказівки, пов’язані з цими критеріями, повинні бути написані простим способом, щоб їх було легко зрозуміти. | Важливий |
| Система повинна містити механізми відновлення у разі збою,<br>яка включає, наприклад, проблеми з послугою хостингу та збої доступу до зовнішніх ресурсів (API). | Бажаний |
| Додаток має враховувати функції доступності, не нав’язуючи умови роботи, які перешкоджають користувачам, які використовують функції читання з екрана та інші. | Важливий |

## 3.2 Розробка архітектури модуля виявлення неправдивих новин

Передбачається, що розроблений додаток буде представляти веб-систему, яку можна використовувати як на настільних комп'ютерах, так і на смартфонах і складається з двох модулів. Перший модуль, який називається «Модуль аналізу

новин», спрямований на підтримку користувача в процесі бічного читання. Цей підхід допомагає користувачам ретельніше оцінювати джерела інформації, аналізуючи їхні характеристики та метадані, які витягуються з посилання на конкретну новину. Модуль виконує роль інструменту, який автоматично виділяє ключові дані з посилання, наданого користувачем. Метою цього є створення відправної точки для подальшого дослідження, що дозволяє користувачеві автономно формувати власну думку щодо правдивості чи достовірності контенту. Другий модуль, «Навчальний модуль», використовує підхід, заснований на теорії щеплення. Його завдання — навчати користувачів розпізнавати неправдиву інформацію шляхом практики. Модуль демонструє приклади новин та просить користувача оцінити їхню достовірність. Цей процес спрямований на підвищення медіаграмотності користувачів і покращення їхніх навичок аналізу інформації.

Навчених і високоточних моделей машинного навчання у вакуумі недостатньо для вирішення проблеми виявлення фейкових новин. Таким чином, ми розробили підтвердження концепції для виявлення фейкових медичних новин, пов’язаних із захворюваннями, які використовують згадані вище навчені моделі машинного навчання. Ці моделі забезпечують високу точність розпізнавання текстових особливостей, що є важливим кроком у боротьбі з дезінформацією.

Наш веб-додаток – це інструмент, який допомагає користувачеві оцінювати достовірність медичної інформації в режимі реального часу. Завдяки інтуїтивно зрозумілому інтерфейсу та швидкому аналізу даних користувач отримує результати, які можуть бути основою для подальших дій, таких як перевірка альтернативних джерел або звернення до спеціалістів.

Першим шаром у моделі є шар BERT. Цей шар відповідає за створення контекстно-залежних вставок слів, що дозволяє краще розуміти значення тексту в конкретному контексті. Перший блок використовує модель BERT для формування якісних текстових репрезентацій, які потім передаються як вхідні дані до другої частини – нейронної мережі, яка забезпечує подальший аналіз і класифікацію

новини. Ця багаторівнева архітектура дозволяє обробляти складні текстові дані з високою точністю, сприяючи виявленню навіть прихованих ознак дезінформації.

На рисунку 3.1 показані детальні параметри кожного шару в моделі.

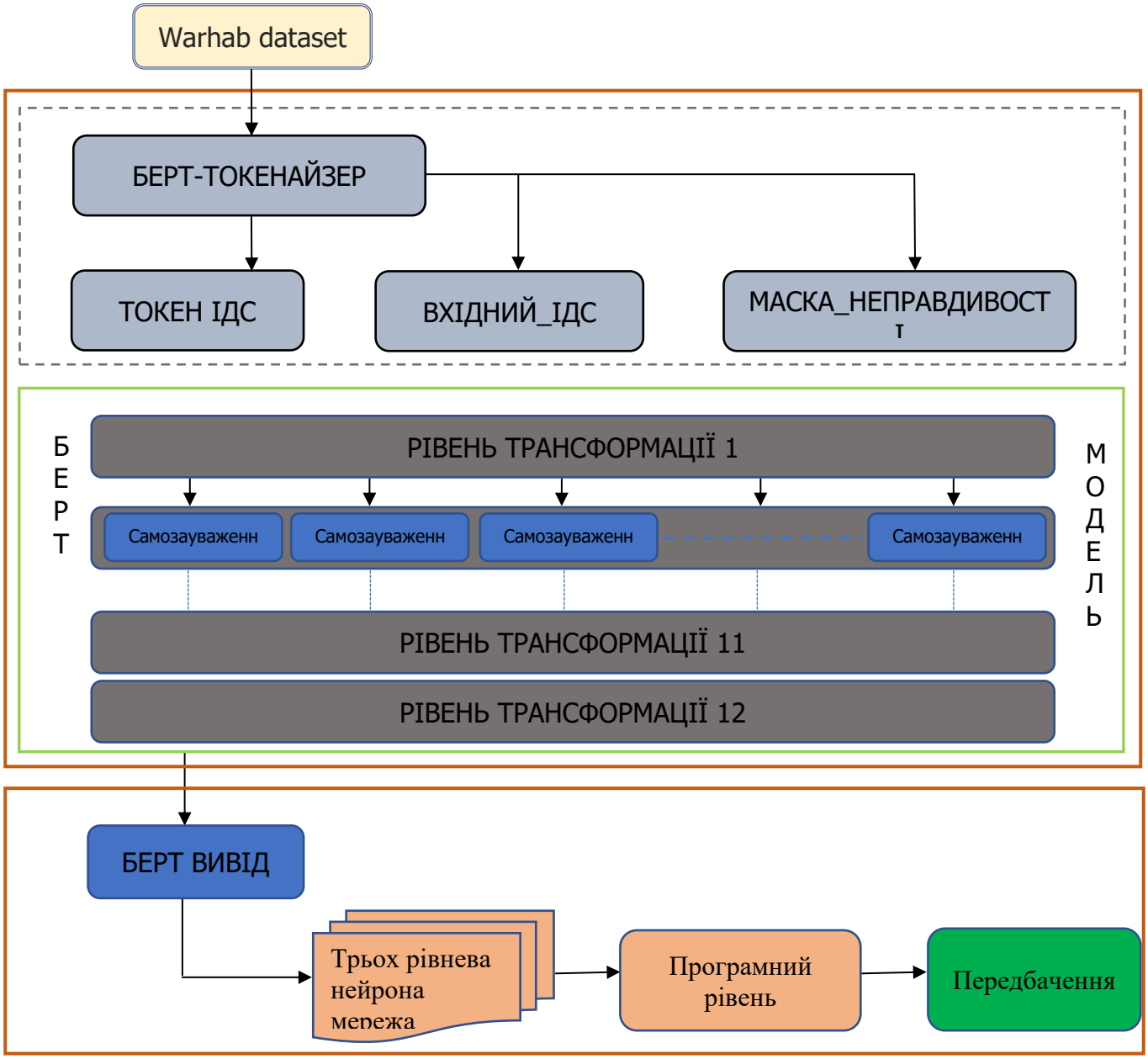



Рисунок 3.1 – Система виявлення неправдивої новин

У наших дослідженнях ми використовуємо базову модель BERT. Таблиця 3.4 показує деталі моделі BERT, використаної в нашому дослідженні. Модель має 12 трансформаторних шарів; кожен рівень виконує механізм самоконтролю і передає його на наступний рівень прямої мережі.

Механізм само уважності є основним будівельним блоком трансформатора в BERT. Кожен шар виводить вектор розміром 768. Коли вбудовування слова проходить через різні шари в блоці трансформатора, воно вивчає контекст усього речення. Вироблені верхні вбудовування повністю контекстуалізовані та враховують усе речення. Ми використовуємо трансформатори та бібліотеку TensorFlow для використання BERT.

Таблиця 3.4 BERT-базова архітектура

| Назва параметра | Значення |
|---|---|
| Модель BERT | BERT в базовому корпусі |
| Кількість кодерів | 12 |
| Кількість голів уваги | 12 |
| Розмір створених вставок/розмір прихованого шару | 768 |
| Загальна кількість параметрів | 110M |
| Розмір введення | 100 |
| Бібліотека, яка використовується для впровадження BERT | трансформери |
| Токенізатор | Токенізатор BERT |

Модель BERT виводить останній прихований стан і вихід пулера. Останній прихований стан - це послідовність прихованих станів на останньому рівні моделі BERT. Об’єднаний вихід - це послідовність прихованих станів маркера CLS, що проходить через лінійний рівень і функцію активації Tanh. Останній прихований стан і вихід Pooler мають розмір (розмір партії, фіксована довжина введення, 768) і (розмір партії, 768), відповідно. Будь-які найпопулярніші вбудовування можна додати до класифікатора, оскільки вони повністю контекстуалізовані. Тим не менш, існує ймовірність, що вони можуть бути локалізовані відповідно до значення конкретної лексеми. Отже, щоб обійти цей артефакт, ми використовуємо

вбудовування CLS (об’єднаний вихід), оскільки він не зосереджується на жодній окремій лексемі в реченні.

Вихід пулера подається як вхід до щільної, повністю пов’язаної нейронної мережі з 3 рівнів, за якою слідує останній рівень сигмоїдної функції активації. Розмір партії, який використовується для навчання, становить 8. Після експериментів з різними атрибутами досягається 99% точності. Таблиця 3.5 показує деталі гіперпараметрів моделі виявлення неправдивої інформації FakeOut.

Таблиця 3.5 - Структура системи

| Назва параметра | Значення |
|---|---|
| Розмір партії | 8 |
| Розмір навчальних даних | 80% |
| Розмір даних перевірки | 10% |
| Розмір тестових даних | 10% |
| Оптимізатор | Адам |
| Функція втрат | Двійкова перехресна ентропія |
| Метрики | Точність, точність і запам'ятовування |

Пропонована система була розроблена з урахуванням специфіки роботи у віртуальних середовищах, що забезпечує її максимальну гнучкість і доступність для користувачів. Основний акцент зроблено на адаптацію до потреб сучасних користувачів, які взаємодіють із системами переважно через інтернет, використовуючи різні пристрої.

Вибір роботи у віртуальних середовищах базувався на аналізі сучасних тенденцій у використанні веб-технологій. Завдяки цьому підходу система може бути доступною як через стаціонарні комп’ютери, так і через мобільні пристрої, такі як смартфони й планшети. Це дозволяє охопити широку аудиторію,

забезпечуючи легкий доступ до сервісів незалежно від місця розташування користувача.

## 3.3 Проектування аналітичного модуля програмної системи виявлення неправдивих новин

Аналітичний модуль FakeOut представляє собою веб-додаток, що дозволяє користувачеві ввести текст, вибрати техніку векторизації (TF-IDF або Count Vectorizer) і модель машинного навчання. Потім він використовує навчену модель для прогнозування достовірності введеного тексту. Оскільки ми вважаємо, що жодна модель машинного навчання не є абсолютно точною та надійною для виявлення фейкових новин, ми також показуємо п’ять найпопулярніших результатів пошуку Google за введеним текстом. Користувачам надається можливість вибрати, погоджуватися вони з оцінкою моделі чи ні. Цей відгук служить подвійним цілям. Це допоможе в безперервному оновленні WarHub, оскільки введений текст буде збережено на сервері та покращить продуктивність моделі. По-друге, у міру того, як збирається все більше і більше даних, позначених людиною, веб-додаток можна розширити, щоб у майбутньому відображати рейтинги достовірності тексту, отримані натовпом.

Інтерфейс програми – це простий графічний інтерфейс, написаний за допомогою HTML і CSS. Верхня частина може бути реалізована за допомогою будь-якого фреймворку.

У відповідності до описаного у попередніх розділах, було використовуємо модель BERT для представлення найефективнішої моделі системи виявлення неправдивих новин. Це п’ятирівнева модель глибокого навчання на основі BERT, яка складається з двох основних будівельних блоків.

Архітектура інструменту клієнт-сервер. Що стосується використовуваних технологій, то були прийняті інструменти, що використовуються для створення веб-систем: HTML, CSS, JavaScript. Хоча перші два використовуються спеціально для створення графічних інтерфейсів на стороні клієнта, мова програмування JavaScript використовується у всьому програмному забезпеченні. Зокрема, фреймворк Vue.js допомагає створювати інтерфейси, полегшуючи розробку динамічних сторінок, які також містять HTML і CSS.

Що стосується сторони сервера, Node.js забезпечує середовище виконання для JavaScript на серверах. Однією з наданих можливостей у цьому сенсі є доступ до зовнішніх API, а також створення нових API - важливих функцій для розробки цього проекту.

На рисунку 3.2 показана схема архітектури та зовнішніх компонентів, що використовуються системою. Окрім повної реалізації навчального модуля, клієнтська сторона програми відповідатиме за відображення даних, які складають модуль аналізу новин. Щоб отримати ці дані, надсилаються запити на сервер, який, у свою чергу, використовує бібліотеки, API та інші платформи для вилучення цих даних, виконує відповідну обробку та забезпечує маршрути доступу до цієї інформації через протокол HTTP.

Важливим обговорюваним моментом у цій архітектурі є інструменти (бібліотеки, API та сайти), які використовуються для отримання даних, що використовуються для допомоги користувачеві у виконанні бокового читання. Ця інформація, яка відображається в модулі аналізу новин, стосується критеріїв, викладених у розділі «Модуль аналізу новин». Ці інструменти будуть розглянуті нижче.

Metascraper – це бібліотека Node.js, яка дає змогу витягувати (вирізати) метадані з он-лайн статей (наприклад, новин, публікацій у блогах тощо). Окрім текстових даних, інструмент також може отримувати зображення, відео та аудіо.

Бібліотека робить це, аналізуючи елементи HTML, присутні на сторінці цікавої статті. Для частини інформації існує набір селекторів, які можна

використовувати для її вилучення. Наприклад, щоб отримати ім’я автора, можна відфільтрувати HTML-елементи, у яких атрибут «class» або атрибут «name» містить термін «author». Коли в одному з цих елементів знайдено непорожнє значення, пошук припиняється і повертаються знайдені дані.

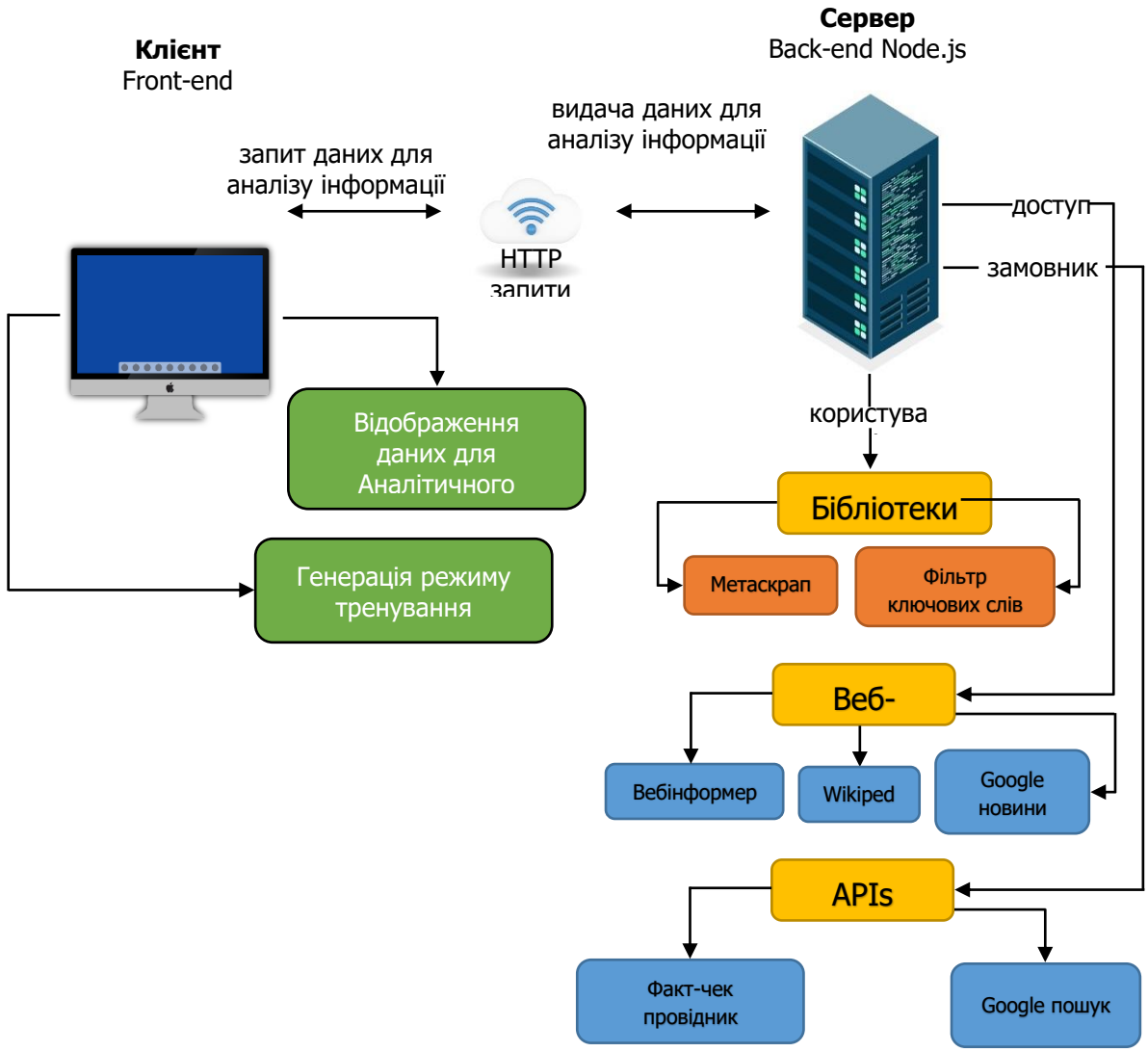



Рисунок 3.2 - Архітектура аналітичного модуля системи

У контексті цього дослідження бібліотека використовувалася для отримання таких даних: автор, назва та дата публікації інформації (новини).

Екстракт ключових слів : це бібліотека Node.js, яка вибирає ключові слова з рядка, що містить фразу. Цей процес здійснюється шляхом видалення стоп-слів. Стоп-слова – це «шумові слова», тобто слова, які не мають особливого значення чи

значення в текстовому виразі. Видалення цих слів із фрази зазвичай використовується для підвищення точності результатів, які повертаються під час пошуку в онлайн-пошукових системах, таких як Google. Деякі приклади стоп-слів в українській мові: «до», «від», «той».

У цьому дослідженні виділення ключових слів було необхідним для здійснення перевірки фактів, пов’язаних з проаналізованими новинами та новинами, схожими на проаналізовані новини, але опублікованими в інших джерелах. Обидва процеси пояснюються нижче.

Google Search API (Google) – цей API, розроблений компанією Google, дає змогу здійснювати необмежений пошук у системах Google. Надані результати містять посилання, описи та веб-сайти. Використання API дуже просте та зрозуміле, але воно також дає змогу налаштовувати такі параметри, як мова, національність і географічне розташування результатів.

На додаток до звичайної пошукової системи Google, API також надає доступ до пошуку за допомогою Google Images (служба пошуку зображень), Google News (агрегатор новин, який індексує новини з різних джерел) і Google Scholar (пошукова система для наукових статей).

Fact Check Explorer – цей інструмент Google допомагає досліджувати журналістів і дослідників, забезпечуючи систему пошуку новин, які вже були перевірені агентствами з перевірки фактів. Таким чином можна, використовуючи термін або фразу, отримати найновіші результати, опубліковані агентствами з перевірки фактів, і таким чином проаналізувати достовірність певного предмета.

Fact Check Explorer містить будь-яку перевірку фактів. Агентство, яке дотримується вказівок Google, які включають, наприклад, надання форматування даних відповідно до стандарту, наявність політики виправлення або механізмів, за допомогою яких користувачі можуть повідомляти про помилки, веб-сайт агентства не може належати політичній організації, а перевірка фактів має бути прозорою. щодо джерел і методів – містять цитати та посилання на першоджерела.

Існує також API, який дозволяє використовувати Fact Check Explorer під час виконання. Використовуючи цей API, можна було отримати та відобразити основні перевірки фактів, пов’язані з новинами, проаналізованими в розробленій системі. Використовуваний пошуковий рядок складається з перших трьох ключових слів у заголовку (отриманих за допомогою засобу вилучення ключових слів ), які об’єднані логічним оператором «I».

Потім знайдені перевірки фактів відображаються в системі разом із відповідними оригінальними посиланнями, дозволяючи користувачеві отримати доступ до веб-сайтів агенцій із перевірки фактів і перевірити аргументи, наведені перевіриними фактами, щоб продемонструвати, що новини неправдиві. Метою надання посилань на перевірку фактів є спонукання користувачів до деякого стороннього читання.

Інформатор веб-сайту, одна з частин інформації, що відображається в системі, стосується особи (фізичної чи юридичної особи), відповідальної за реєстрацію домену, на якому розміщено проаналізовану новину. Коли доменне ім’я купується через певного постачальника, інформація про особу, відповідальну за покупку, передається до ICANN (Інтернет-корпорації з присвоєння імен і номерів), некомерційної організації, яка, серед інших обов’язків, керує протоколом DNS (Система доменних імен). Деякі дані в цих доменах доступні для загального доступу через WHOIS, протокол, яким також керує ICANN. Варто зазначити, що деякі провайдери пропонують послугу збереження конфіденційності даних власника субдомену. Для цього вони подають власні дані в ICANN замість даних своїх клієнтів.

Website Informer – це система, яка отримує доступ до WHOIS і дозволяє відображати дані публічного субдомену. У проекті, описаному тут, ця платформа використовувалася для отримання та відображення, коли це було відкрито, імені особи, відповідальної за реєстрацію субдомену.

Вікіпедія. Під час аналізу історії важливо знати ЗМІ, відповідальне за поширення новин. Для цього система виводить інформацію, взяту з Вікіпедії, про джерело, в якому опублікована новина. Ця функція реалізується в два етапи:

На першому етапі здійснюється пошук за допомогою API веб-пошуку Google за URL-адресою сайту, на якому опубліковано аналізовану новину. Для цього пошуку застосовано такий фільтр: лише результати з домену «en.wikipedia.org» (підрозділ Вікіпедії португальською мовою). Таким чином можна знайти посилання на потрібну статтю у Вікіпедії.

Далі здійснюється HTTP-запит до отриманого посилання на статтю у Вікіпедії. У результаті отримується HTML-код сторінки, з якого витягується текст першого абзацу за допомогою аналізу класів і тегів HTML-коду.

Новини Google, окрім збору новин із різних джерел, новини Google дозволяють здійснювати пошук у базі даних новин. Інша можливість полягає в організації новин відповідно до вподобань користувача.

У цьому дослідженні цей механізм використовувався для відображення інших новин, пов’язаних із матеріалом, який аналізується. Це ще одна частина інформації, яка спрямована на покращення бокового читання, зрештою, вона надає ряд пов’язаних новин і дозволяє користувачеві мати ширший погляд на тему. Пошук новин здійснювався за допомогою Google Search API. Використовуваний пошуковий рядок складається з трьох ключових слів у заголовку новини (отриманих Keywords Extractor ), об’єднаних логічним оператором «I». Пошук все ще обмежується новинними матеріалами португальською мовою, які були опубліковані на бразильських веб-сайтах. Система відображає перші 5 знайдених новин.

Цей модуль аналізу запускається, коли користувач вставляє посилання на новину на домашній сторінці програми. Інформація, висвітлена в модулі, має на меті заохотити та стимулювати користувача зрозуміти контекст, у який вставляються аналізовані новини, звернутися до інших джерел і виявити можливі

упередження в засобі масової інформації, у якому публікується інформація, таким чином заохочуючи практику латерального читання.

Передбачене та правильне функціонування системи залежить від вставки текстових посилань на новини, опублікованих на веб-сайтах чи блогах, які мають такі характеристики, як: назва, дата публікації та текст. Якщо ввести будь-який інший тип посилання, інформація, яка відображається системою, буде неможливою, і, як наслідок, інструмент не зможе виконувати свою мету: допомагати в процесі бічного читання. Крім того, повідомлення про помилку відображається, коли користувач вводить у відповідне поле рядок, який не має формату URL-адреси.

Варто зауважити, що існують ситуації, коли система не може отримати необхідну інформацію – або через те, що інформації насправді немає в новині, або через технічну неможливість інструментів, прийнятих у реалізації. У таких ситуаціях відображаються повідомлення, які пояснюють користувачеві, що інформацію отримати неможливо.

У зв’язку з цим варто виділити дві ситуації, коли неможливо знайти пов’язані новини або перевірку фактів, пов’язані з новиною, що аналізується. Якщо пов’язаних новин не знайдено, відображається повідомлення, яке вказує на відсутність результатів і пояснює, що це може статися у випадках, коли новина, яка аналізується, не була широко поширена в новинних виданнях – зрештою, інструмент, який використовується для цього пошуку, індексатор Google для новин, який збирає історії з багатьох джерел новин.

Більше того, якщо пошук не знаходить помилкових перевірок, система сповіщає користувача про те, що проаналізовані новини можуть бути правдивими. Цей висновок ґрунтується на тому, що, оскільки не виявлено публікацій, які спростовують новину, вона може бути правдивою. Варто зауважити, що відображене повідомлення не має на меті надати користувачеві впевненості, а скоріше служити основою, разом з іншою відображеною інформацією, для формування власного висновку.

Новий модуль аналізу також надає пояснення для кожної виділеної в ньому частини інформації, щоб користувач міг зрозуміти, як кожну частину інформації можна використовувати в процесі розуміння новин. Щоб отримати доступ до цих пояснень, клацніть піктограму зі знаком питання («?»), розташовану поруч із заголовком кожного поля з інформацією.

Навчальний модуль надає користувачам можливість відпрацювати свої навички визначення фейкових новин і, перш за все, як спосіб застосування концепції теорії щеплення. Крім того, щоб забезпечити ігровий досвід, модуль також має елементи ігрового процесу, такі як підрахунок очок, візуальний зворотній зв’язок і звукові ефекти.

За один раунд використання модуля користувачеві буде представлено 5 прикладів реальних новин. Щоб один і той же користувач міг використовувати його кілька разів, реєструється 20 різних новин, і п’ять з них відображаються випадковим чином у кожному раунді.

Для кожної новини буде виділена така інформація:

- ЗМІ;
- назва;
- дата публікації;
- текст;
- посилання на місце публікації (за наявності).

Варто зазначити, що деякі новини, відтворені в модулі, були опубліковані в додатках для обміну повідомленнями (переважно WhatsApp і Telegram), а тому не мають посилання. Для інших новин користувач може натиснути кнопку та отримати доступ до модуля аналізу новин, маючи таким чином доступ до інформації, яка може допомогти йому зробити висновок щодо правдивості представленої історії.

Якщо аналіз правдивості новини зроблений користувачем правильно, він отримає збільшення балів на 10 балів. Ця оцінка не буде збільшена, якщо відповідь буде неправильною. У таких випадках користувачеві буде показано коротке

пояснення, в якому будуть вказані причини та аргументи, чому правдивість новини не відповідає відповіді.

В кінці раунду користувачеві буде показано його рахунок. Максимальна загальна сума балів, яку можна отримати, становить 50.

Цей механізм спрямований на підвищення обізнаності користувачів щодо розпізнавання дезінформації та сприяє розвитку критичного мислення через інтерактивний формат навчання.

## 3.4 Висновки до 3-го розділу

У третьому розділі було визначено ключові аспекти, що забезпечують її функціональність, ефективність та відповідність потребам користувачів. Чітко сформульовані функціональні та нефункціональні вимоги стали основою для створення модулів аналізу та навчання, кожен із яких виконує важливу роль у досягненні загальної мети – підвищення медіаграмотності та запобігання поширенню дезінформації.

Функціональні вимоги зосереджені на забезпеченні основних можливостей системи, таких як аналіз текстових даних, інтеграція з API, генерація звітів та надання користувачам інструментів для аналізу достовірності новин. Натомість нефункціональні вимоги спрямовані на покращення якості роботи системи, зокрема забезпечення продуктивності, масштабованості, безпеки, зручності використання та сумісності.

Архітектура системи передбачає використання моделі машинного навчання BERT, що дозволяє досягти високої точності виявлення фейкових новин завдяки глибокому аналізу тексту та контексту. Інтеграція зовнішніх інструментів, таких як Metascraper, Google Search API та Fact Check Explorer, забезпечує доступ до

додаткових даних і підтримує практику бокового читання, сприяючи глибшому розумінню контенту.

Аналітичний модуль FakeOut реалізує підхід, що поєднує автоматизовану перевірку новин із залученням користувачів до процесу оцінки достовірності. Зворотній зв’язок від користувачів сприяє покращенню моделі через постійне оновлення бази даних і точності алгоритмів.

Навчальний модуль забезпечує інтерактивний формат підвищення обізнаності щодо дезінформації через ігрові елементи, такі як підрахунок балів, пояснення відповідей і демонстрація реальних прикладів новин. Цей підхід сприяє розвитку критичного мислення та навичок медіаграмотності.

Загалом, проектовані модулі та архітектурні рішення забезпечують комплексний підхід до виявлення фейкових новин, поєднуючи сучасні технології, інтерактивне навчання та підтримку користувачів у боротьбі з дезінформацією. Така система є не лише корисним інструментом, а й ефективним засобом формування критичного ставлення до медіаконтенту.

Аналітичний модуль FakeOut представляє собою веб-додаток, що дозволяє користувачеві ввести текст, вибрати техніку векторизації (TF-IDF або Count Vectorizer) і модель машинного навчання. Потім він використовує навчену модель для прогнозування достовірності введеного тексту.

# 4 ОЦІНКА АНАЛІТИЧНОГО МОДУЛЯ

## 4.1 Оцінка використання аналітичного модуля

Після проектування та розробки аналітичного модуля можлива теоретична та практична перевірка потенційними користувачами. Для цього доцільно розробити та затвердити протокол валідації, а також спостереження за використанням спостерігачем. У цьому розділі представлені як методичні рішення, так і результати цього процесу.

Можливий процес перевірки та оцінки може проходити у два етапи: перший із широким загалом, а другий – з користувачами старших вікових груп. Методологія, що керує валідацією, є переважно кількісною та базується на зборі, обробці та статистичному аналізі інформації за допомогою Анкети досвіду користувача. Незважаючи на те, що дослідження переважно кількісне, воно ткож може включати аспекти якісної оцінки, беручи до уваги виступи та доповіді учасників дослідження, а на другому етапі також можуть бути особи, яких можна визначити як спостерігачі, тобто вони можуть допомагати членам своїх сімей (тестерам) використовувати інструмент і в цьому супроводі записувати спостереження за своїм сприйняттям.

Анкета досвіду користувача. В якості механізму доцільно використовувати структурований опитувальник UEQ (User Experience Questionnaire). Цей інструмент є інструментом оцінки програмного забезпечення, що складається з 26 елементів, що вимірюють як класичні аспекти зручності використання, так і характеристики взаємодії з користувачем. Розуміння цих аспектів (характеристик) можна узагальнити в 6 показниках, представлених і пояснених у списку нижче.

Привабливість – загальне враження від товару. Подобається або не подобається продукт користувачам?

Чіткість – чи легко користувачам ознайомитися з продуктом? Чи легко навчитися користуватися продуктом?

Ефективність – чи можуть користувачі виконувати свої дії в інструменті без зайвих зусиль?

Надійність – чи відчуває користувач контроль під час взаємодії з продуктом?

Стимуляція (мотиваційність) – чи є використання продукту захоплюючим і стимулюючим?

Інновація (новизна) – чи є продукт інноваційним і креативним? Чи захоплює продукт інтерес користувачів?

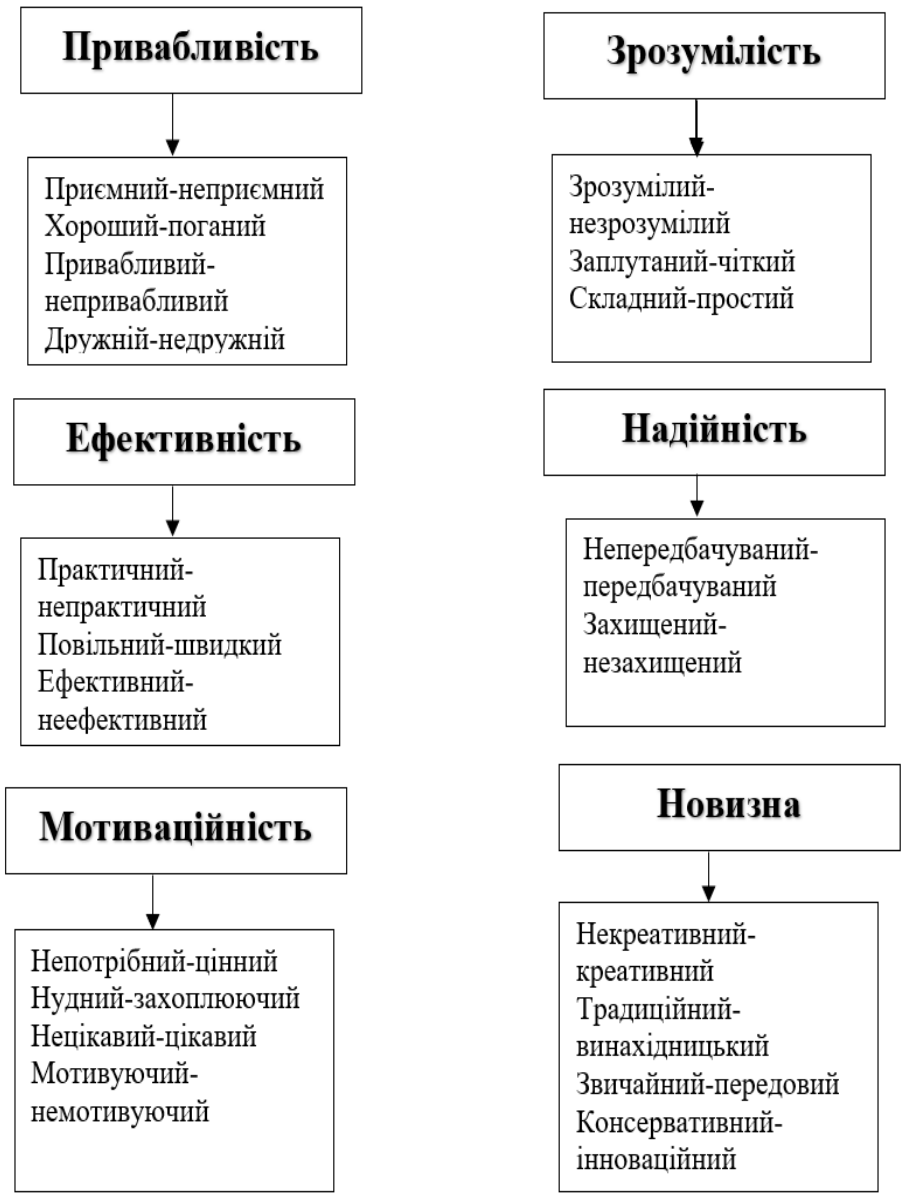



Рисунок 4.1 – Список критеріїв та пов’язаних із ними елементів

Шкала Привабливості в опитувальнику складається з 6 пунктів, а інші – з 4 пунктів. Повну анкету, що містить 26 пунктів, можна знайти в таблиці 4.1. На рисунку 3.1 показано діаграму, що пояснює пункти, які складають оцінку для кожного з критеріїв.

Відповідь на анкету здійснюється шляхом присвоєння балів кожному пункту. Оцінки можуть змінюватися за шкалою Лайкерта із 7 пунктів – оцінки в діапазоні від 1 до 7. Кожній оцінці присвоюється вага від -3 (для найбільш негативної відповіді: 1) до +3 (для найбільш позитивної відповіді: 7) . Такі фактори анкети, як надійність (забезпечення узгодженості використовуваних шкал) і валідність (забезпечення того, що шкали дійсно оцінюють те, для чого вони призначені), вже досліджувалися в 11 тестах на юзабіліті за участю 144 учасників і за допомогою онлайн-опитувальника з 722 учасниками [32].

Таблиця 4.1 – Анкета перевірки взаємодії з користувачем

| **Перелік** | **Оцінка (1, 2, 3, 4, 5, 6, 7)** |
|---|---|
| Дратівливий | Приємний |
| Незрозумілий | Зрозумілий |
| Здатний | Тупий |
| Творчий | Нетворчий |
| Легко вчитися | Важко вчитися |
| Цінний | Неповноцінний |
| Нудний | Захоплюючий |
| Нецікавий | Цікавий |
| Непередбачуваний | Передбачуваний |
| Швидкий | Повільний |
| Винахідливий | Звичайний |
| Обструктивний | Підтримуючий |
| Добре | Поганий |

Продовження таблиця 4.1

| **Перелік** | **Оцінка (1, 2, 3, 4, 5, 6, 7)** |
|---|---|
| Неприємний | Приємний |
| Складний | Легкий |
| Звичайний | Передовий |
| Безпечний | Незахищений |
| Мотивуючий | Демотивуючий |
| Відповідає очікуванням | Не відповідає очікуванням |
| Неефективний | Ефективний |
| Ясний | Збиває з пантелику |
| Непрактичний | Практичний |
| Організований | Хаотичний |
| Привабливий | Непривабливий |
| Дружній | Недоброзичливий |
| Консервативний | Інноваційний |

У цьому дослідженні перевірка передбачала надсилання посилання, за яким можна було отримати доступ до програми, разом із коротким вступом до проекту та анкетою UEQ. Після того, як інформацію було зібрано, ми приступили до її аналізу та робили можливі висновки – як це передбачено прийнятим документом.

## 4.2 Метрика оцінки та валідації програмного продукту

Спостереження за користувачами старших вікових груп є одним із факторів, що доцільно запроваджувати на другому етапі валідації інструменту. Цей підхід враховує особливості цієї категорії користувачів, які можуть мати специфічні

потреби та труднощі у використанні сучасних цифрових технологій. Зважаючи на те, що більшість представників старших вікових груп народилися в епоху, коли технологічні інструменти ще не були доступні, їхній досвід користуванні такими системами може бути обмеженим.

У цьому контексті процес валідації передбачає участь спостерігачів, які можуть не лише аналізувати дії учасників, але й допомагати їм під час використання системи. Спостерігачами можуть бути члени сім'ї учасників (наприклад, діти, онуки або інші родичі), які підтримують літніх користувачів і записують свої спостереження щодо їхнього досвіду роботи з інструментом.

Під час проведення спостереження рекомендується використовувати чітко сформульовані питання, які допомагають спостерігачам документувати ключові моменти взаємодії літніх користувачів із системою. Нижче наведено приклади таких питань.

Дії, реакції та висловлювання – це питання спрямоване на фіксацію емоційної реакції користувачів, їхньої залученості до процесу використання інструменту, а також будь-яких незвичайних або несподіваних дій, які можуть вказувати на потенційні проблеми чи позитивні сторони інтерфейсу. Приклад питання: «Які цікаві дії, реакції або висловлювання ви спостерігали під час використання інструменту вашим батьком, матір'ю, дідусем, дядьком, свекром чи іншими родичами? ».

Труднощі – це питання дозволяє виявити конкретні аспекти інструменту, які можуть бути складними для літніх користувачів. Наприклад, це можуть бути дрібні шрифти, незрозумілі інструкції, складний процес навігації чи технічні перешкоди, які заважають ефективному використанню. Приклад питань: «Чи помічали ви якісь труднощі, які виникали під час роботи з інструментом? Якщо так, то в чому полягали ці труднощі?».

Фактори, що полегшують використання – це питання спрямоване на визначення позитивних сторін інструменту, таких як зручні кнопки, логічна структура інтерфейсу чи корисні підказки, які допомогли користувачам краще

зрозуміти, як працює система. Приклади питань: «Чи були аспекти, які зробили використання інструменту більш зрозумілим і легким? Якщо так, то які саме?».

Повторне використання – це питання допомагає оцінити рівень задоволеності користувачів системою. Якщо користувачі зацікавлені в повторному використанні, це свідчить про те, що інструмент відповідає їхнім очікуванням і приносить користь. У разі відсутності бажання важливо з'ясувати причини, щоб покращити функціональність та зручність системи. Приклад питань: «Чи виявляли користувачі бажання знову скористатися інструментом? Якщо так, то які причини вони наводили або які ви помітили? Якщо ні, то які фактори цьому сприяли?».

Для забезпечення конфіденційності учасників і спостерігачів усі дані під час аналізу будуть знеособлені. Учасників ідентифікуватимуть за допомогою номерів, таких як Учасник 1, Учасник 2 тощо. Це дозволить уникнути можливості ідентифікації особи та забезпечить чесність і об'єктивність результатів

Процес перевірки відбувався за допомогою форми, яка складалася спочатку з питань про профілі учасників, а потім із питань, запропонованих інструментом UEQ [31]. для оцінки зручності використання. Цей етап життєвого циклу проекту розділяється на два етапи.

Перший етап – широке тестування.

– на цьому етапі опитування відкривається для широкої аудиторії, включаючи користувачів різного віку та рівня цифрової компетенції;

– мета цього етапу – отримати загальні відгуки та виявити основні проблеми у функціональності системи.

Другий етап – фокус на літніх користувачах.

– цей етап зосереджений на вивченні специфічних труднощів, які можуть виникати у користувачів старших вікових груп;

– метою є адаптація системи для цієї категорії, враховуючи їхні потреби та можливості.

В обох випадках оцінювання може відбуватись онлайн, а посилання для доступу до системи та форма оцінювання можуть бути розповсюджені через список

електронної пошти, наприклад університету і повідомлення через соціальні мережі (Discord, WhatsApp, Viber тощо). Користувачі також можуть отримати посібник з інструкціями щодо користування додатком із поясненнями використання обох модулів додатку.

## 4.3 Висновки до 4-го розділу

У даному розділі розглянуто теоретичні та практичні аспекти перевірки розробленої системи потенційними користувачами. Для забезпечення достовірності та точності результатів необхідно розробити детальний протокол валідації, який буде включати як етапи тестування, так і умови проведення спостережень. Важливу роль у цьому процесі відіграє безпосередня взаємодія користувачів із системою, що дозволяє оцінити її зручність, ефективність і відповідність поставленим завданням.

Процес перевірки та оцінки ефективності системи передбачає два ключових етапи. На першому етапі тестування залучається широкий загал користувачів різних вікових і соціальних категорій. Це дозволяє зібрати різноманітні відгуки, виявити загальні закономірності у використанні системи та визначити основні аспекти, які потребують удосконалення. Другий етап зосереджується на старших вікових групах, оскільки їхні потреби та вимоги до інтерфейсу можуть суттєво відрізнятися. Особливу увагу приділяють спрощенню взаємодії з системою, забезпеченню зрозумілості функцій та адаптації до фізичних і когнітивних особливостей цієї категорії користувачів.

Важливим компонентом перевірки є надсилання додаткової анкети на другому етапі тестування. Ця анкета має на меті зібрати додаткові дані щодо специфічних випадків, коли валідацію проводять спостерігачі або помічники користувачів старшого віку. Це може бути важливим у ситуаціях, коли людина

стикається з труднощами у використанні системи через обмеження фізичних чи когнітивних можливостей. Спостерігачі виконують роль посередників, допомагаючи коректно виконувати завдання та надаючи зворотний зв’язок щодо роботи системи.

Таким чином, запропонований підхід до перевірки системи забезпечує її всебічне тестування, враховуючи як загальні сценарії використання, так і специфічні потреби окремих категорій користувачів. Валідація дозволяє не лише оцінити поточний рівень ефективності системи, але й окреслити напрямки для її подальшого вдосконалення.

# ВИСНОВКИ

Отже, в результаті проведеного дослідження можна зробити наступні висновки.

У сучасному світі проблема поширення неправдивої (фейкової) інформації набуває загрозливих масштабів, впливаючи на політичні, економічні, соціальні процеси та навіть на індивідуальні рішення. Зокрема, дезінформація в соціальних мережах може маніпулювати громадською думкою, створюючи ілюзію правдивості завдяки швидкому поширенню та використанню ботів і підроблених акаунтів. Вирішення цієї проблеми потребує нових підходів, що поєднують сучасні технології та методи, адаптовані до специфіки сучасного інформаційного середовища..

Щоб вирішити проблему відсутності належним чином позначених наборів даних для виявлення фейкових новин, було представлено модуль, який передбачає вміст нового комплексного набору даних WarHub з останніми фактами та міфами про військові дії на території України.

У першому розділі було проведено аналіз предметної області та методів виявлення неправдивих новин. Виділено основні підходи: лінгвістичний аналіз, соціально-контекстуальні методи, перевірка фактів і машинне навчання. Встановлено, що найефективнішим підходом є інтеграція цих методів у вигляді гібридного підходу, який забезпечує комплексний аналіз. Визначено завдання для подальшого вдосконалення системи FakeOut із використанням сучасних технологій.

У другому розділі розглянуто процес проектування програмних модулів та методи попередньої обробки даних. Було створено інструмент WarHub для збору та аналізу даних і визначено структуру набору даних, а також запропоновано модель на основі BERT, яка досягає високої точності у класифікації неправдивих новин. Розроблений алгоритм враховує поведінкові особливості користувачів і вплив штучних елементів, таких як боти, що дозволяє ефективно ідентифікувати дезінформацію.

У третьому розділі було визначено функціональні та нефункціональні вимоги до програмної системи, розроблено її архітектуру та реалізовано два основних модулі: аналітичний та навчальний. Важливим аспектом є використання сучасних інструментів, таких як BERT, для інтеграції аналізу тексту та поведінки користувачів. Представлений модуль навчання дозволяє підвищити обізнаність користувачів щодо дезінформації та розвивати їх критичне мислення.

Аналітичний модуль FakeOut представляє собою веб-додаток, що дозволяє користувачеві ввести текст, вибрати техніку векторизації (TF-IDF або Count Vectorizer) і модель машинного навчання. Потім він використовує навчену модель для прогнозування достовірності введеного тексту.

У четвертому розділі здійснено теоретичне обґрунтування та описано можливу теоретичну та практичну перевірку потенційними користувачами. Для цього доцільно розробити та затвердити протокол валідації, а також спостереження за використанням спостерігачем. У цьому розділі представлені як методичні рішення, так і результати цього процесу.

Можливий процес перевірки та оцінки може проходити у два етапи: перший із широким загалом, а другий – з користувачами старших вікових груп.

Практичне значення отриманих результатів. У даній кваліфікаційній магістерській роботі пропонується удосконалення методів виявлення неправдивих новин, шляхом введення класифікаторів, а також удосконалення методу проектування модулів для знаходження та відображення неправдивих новин потенційним використанням веб-додатку.

В подальшому результати даного дослідження можуть використовуватись при боротьбі із розповсюдженням фейкових (неправдивих) новин у різноманітних медійних засобах, зокрема у соціальних мережах, що особливо є актуальним в сучасних умовах ведення бойових та військових дій на території України при боротьбі з ІПСО.

# ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ


1. Chunyuan Yuan, Qianwen Ma, Wei Zhou, Jizhong Han, Songlin Hu. Early Detection of Fake News by Utilizing the Credibility of News, Publishers, and Users Based on Weakly Supervised Learning, *arXiv*, 2020. URL: https://arxiv.org/abs/2012.04233.
2. Xinyi Zhou, Atishay Jain, Vir V. Phoha, Reza Zafarani. Fake News Early Detection: A Theory-driven Model, *Digital Threats: Research and Practice*, pp. 1-25, 2020. DOI: 10.1145/3377478.
3. Media Literacy Ireland (2019) *Be media smart.* Available at: https://www.bemediasmart.ie/ (дата звернення 10.10.2024).
4. NewseumEd (2020) *Is this story share-worthy?* Available at: https://newseumed.org/tools/lesson-plan/story-share-worthy (Accessed: 10 Oktober 2024).
5. On the Media (2016) *Breaking news comsumer handbook: fake news edition.* Available at: https://www.wnycstudios.org/podcasts/otm/segments/breaking-news-consumer-handbook-fake-news-edition (дата звернення 20.09.2024).
6. What isv *fake news?* Available T: https://www.rappler.com/technology/social-media/189656-fake-news-explainer (дата звернення 20.09.2024).
7. Rubin, V., Chen, Y., and Conroy, N. Deception detection for news: three types of fakes, *Proceedings of the Association for Information Science & Technology*, 52 (1), pp.1-4.
8. Salgado, C. *Lies in the social network.* Available at: http://cristosalgado.com/ (Accessed: 23 September 2024).
9. Spencer, J. *5C's of critical consuming.* Available at: https://www.youtube.com/watch?v=xf8mjbVRqao (Accessed: 10 Oktober 2024).

10. The UNESCO Courier (2017) *Fake news: sound bites on a burning topic*. Available at: https://en.unesco.org/courier/july-september-2017/fake-news-sound-bites-burning-topic (Accessed: 24 September 2024).
11. Analysis of training methods and neural network tools for fake news detection. June 2023. Cybersecurity Education Science Technique. 4(20):20-34.
12. Oxford Word of the Year 2016|Oxford Languages. 16 June 2020. Available online: https://languages.oup.com/word-of-the-year/2016/ (accessed on 10 September 2023).
13. Analysis of training methods and neural network tools for fake news detection. June 2023. Cybersecurity Education Science Technique. 4(20): pp. 20-34.
14. M. H. Goldani, S. Momtazi, and R. Safabakhsh. Detecting fake news with capsule neural networks, Appl. Soft Comput., vol. 101, p. 106991, 2021.
15. Wang J, Zhu Z, Liu C, Li R, Wu X (2024) LLM-Enhanced multimodal detection of fake news. PLoS ONE 19(10): e0312240. DOI: 10.1371/journal.pone.0312240
16. Soga K, Yoshida S, Muneyasu M (2024) Graph-Based Interpretability for Fake News Detection through Topic- and Propagation-Aware Visualization. Computation 12(4): p. 82. DOI: 10.3390/computation12040082
17. Zhang C, Gupta A, Kauten C, Deokar A.V, Qin X (2019) Detecting fake news for reducing misinformation risks using analytics approaches. European Journal of Operational Research, 279(3): pp. 1036–1052. DOI: 10.1016/j.ejor.2019.02.030
18. Zhou X, Zafarani R (2020) A Survey of Fake News: Fundamental Theories, Detection Methods, and Challenges. ACM Computing Surveys 53(5): 1-40. DOI: 10.1145/3395046
19. Jwa H, Oh D, Park K, Kang J, Lim H (2019) ExBAKE: Automatic fake news detection model based on bidirectional encoder representations from transformers (BERT). Applied Sciences, 9(19): 4062. DOI: 10.3390/app9194062
20. Bondielli A, Marcelloni F (2019) A survey on fake news and rumour detection techniques. Information Sciences, 497: 38–55. DOI: 10.1016/j.ins.2019.05.035

21. Ahmed H, Traore I, Saad S (2018) Detecting Opinion Spams and Fake News Using Text Classification. Security and Privacy 1(1): e9. DOI: 10.1002/spy2.9
22. Kaliyar R.K., Goswami A., Narang P. (2020) DeepFakE: Improving fake news detection using tensor decomposition-based deep neural network. Journal of Supercomputing, 77(3): pp. 1015–1037. DOI: 10.1007/s11227-020-03284-3
23. Huang Y.F., Chen P.H. (2020) Fake news detection using an ensemble learning model based on Self-Adaptive Harmony Search algorithms. Expert Systems with Applications, 159, pp. 113-120. DOI: 10.1016/j.eswa.2020.113120
24. Український тиждень. Що таке ІПСО, чому важливо це знати і які операції зараз проводить Росія проти України. URL: https://tyzhden.ua/shcho-take-ipso-chomu-vazhlyvo-tse-znaty-i-iaki-operatsii-zaraz-provodyt-rosiia-proty-ukrainy/ (дата звернення 9.11.2024)
25. Jawahar, G.; Sagot, B.; Seddah, D. What Does BERT Learn about the Structure of Language? In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019.
26. Guarasci, R.; Silvestri, S.; De Pietro, G.; Fujita, H.; Esposito, M. BERT syntactic transfer: A computational experiment on Italian, French and English languages. *Comput. Speech Lang.* 2022, *71*, 101261.
27. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2021.
28. Weiss, K.H.; Khoshgoftaar, T.M.; Wang, D. A survey of transfer learning. *J. Big Data* 2020, *3*, 9.
29. Hamilton, W.L. *Graph Representation Learning*; Springer: Berlin/Heidelberg, Germany, 2020.
30. Shu, K.; Sliva, A.; Wang, S.; Tang, J.; Liu, H. Fake News Detection on Social Media. SIGKDD Explor. 2020.
31. User Experience Questionnaire. URL: https://www.ueq-online.org/.

32. Laugwitz, B., Held, T., Schrepp, M. (2008). Construction and Evaluation of a User Experience Questionnaire. In: Holzinger, A. (eds) HCI and Usability for Education and Work. USAB 2008. Lecture Notes in Computer Science, vol 5298. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-89350-9_6.
33. Salem, F.K.A.; Feel, R.A.; Elbassuoni, S.; Jaber, M.; Farah, M. FA-KES: A Fake News Dataset around the Syrian War. In Proceedings of the International AAAI Conference on Web and Social Media, Münich, Germany, 11–14 June 2019; Volume 13, pp. 573–582.
34. Ahmed, H.; Traore, I.; Saad, S. Detecting opinion spams and fake news using text classification. Secur. Priv. 2017, 1, e9.
35. Koirala, A. COVID-19 Fake News Dataset; Mendeley Data: Amsterdam, The Netherlands, 2021; Volume 1.
36. Disinformation and 7 Common Forms of Information Disorder (2022) https://commonslibrary.org/disinformation-and-7-common-forms-of-information-disorder/
37. Devrim Unay, Pınar Karagoz, Gozde Unay, Alp Aker. MONITOR: A Multimodal Fusion Framework to Assess Message Veracity in Social Networks, *arXiv*, 2021. URL: https://arxiv.org/abs/2109.02271
38. Shu, Kai, et al. "Fake News Detection in Social Networks via Crowd Signals." Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2017, DOI: 10.1145/3097983.3098054.
39. Vosoughi, Soroush, Deb Roy, and Sinan Aral. "The spread of true and false news online." Science, vol. 359, no. 6380, 2018, pp. 1146-1151, DOI: 10.1126/science.aap9559.
40. Zhou, Xiaoyi, and Reza Zafarani. "Fake News: A Survey of Research, Detection Methods, and Opportunities." ACM Computing Surveys, vol. 53, no. 5, 2021, DOI: 10.1145/3395046.
41. Варук В.К., Форкун Ю.В., Аналіз програмних методів та засобів виявлення фейкових новин, APKN-2024 CorpusPaper, 2024, pp.73-78.

# ДОДАТОК А

(обов'язковий)

## КОПІЇ НАУКОВИХ ПУБЛІКАЦІЙ

*Актуальні проблеми комп'ютерних наук*

УДК 004.4

Варук В.К., Форкун Ю.В.

*Хмельницький національний університет*

### АНАЛІЗ ПРОГРАМНИХ МЕТОДІВ ТА ЗАСОБІВ ВИЯВЛЕННЯ ФЕЙКОВИХ НОВИН

*Розглянуто задачу виявлення фейкових новин, а також програмних методів та засобів, що дозволяють здійснювати правильне та чітке здійснення виявлення та знешкодження неправдивої інформації, зокрема у соціальних мережах. Здійснено аналіз методів та засобів виявлення неправдивих новин.*

*The article considers the task of detecting fake news, as well as programmatic methods and means that allow for correct and accurate detection and neutralisation of false information, in particular, in social media. The author analyses the methods and means of detecting fake news.*

Фейкова або неправдива інформація, в тому числі і новини постійно були критичною і складною проблемою в інформаційному медійному просторі. Поширення неправдивих новин викликає хвилювання та глибоке занепокоєння, особливо у сфері медичної інформації, що може мати небезпечні та потенційно смертельні наслідки. В статтях [1,2] дослідники запропонували модель для виявлення фейкових новин, що мають різну довжину новинних тверджень, використовуючи різні варіації вбудовування слів.

Зважаючи на масштабні обсяги дезінформації в Інтернеті, вкрай важливо боротися з фейковими новинами, зокрема медичними та пов'язаними з військовими подіями в Україні. Використання методів модульного проектування для допомоги у виявленні фейкових новин, включаючи медичні й військові, є актуальним завданням, особливо в умовах бойових дій на території України, що робить це питання важливим для дослідження. На сьогоднішній день існують методи виявлення неправдивих новин із використанням ознак, що є однонаправленими (одномодальними), а при злитті направленостей (модальностей), втрачають свою актуальність та достовірність.

Загалом, інформація подається у вигляді простого тексту, тому початкові методи переважно виявляють неправдиву інформацію шляхом вилучення текстових ознак з текстового контенту або вилучення відповідних однонаправлених (одномодальних) ознак з інших однонаправлених даних. З розвитком соціальних мереж форма інформації змінилася з простого тексту на мультимедійну. Більшість існуючої інформації знаходиться у формі мультимедіа. У той же час поєднання особливостей різних модальностей може ефективно покращити продуктивність

моделі. Тому останні дослідження в галузі виявлення неправдивої інформації загалом побудовані із використанням мультимодальних методів [3].

На сьогоднішній день виділяють такі методи для виявлення неправдивих (фейкових) новин:

- на основі контенту;
- заснований на знаннях;
- ручна перевірка фактів;
- краудсорсинг;
- автоматична перевірка фактів;
- лінгвістичні шаблони та текстові ознаки;
- візуальна основа;
- на основі соціального контексту;
- на основі мережі;
- темпоральна поведінка;
- на основі достовірності.

Метод виявлення фейкових новин на основі контенту [5] спрямований на виявлення фейкових новин шляхом аналізу контенту статті, тобто тексту або зображення, або і того, і іншого в новинній статті. Для автоматичного виявлення фейкових новин дослідники часто покладаються на приховані, або створені вручну ознаки контенту.

Підходи, засновані на знаннях [6], використовують метод перевірки фактів, в якому дане твердження порівнюється із зовнішніми джерелами для перевірки достовірності даного твердження. Існуючі методи перевірки фактів можна розділити на ручну (із залученням експертів або за допомогою краудсорсингу) та автоматичну перевірку фактів.

Ручну перевірку фактів можна умовно поділити на перевірку фактів за допомогою експертів та перевірку фактів за допомогою краудсорсингу [7]. Експертна перевірка: Експертні методи використовують експертно-орієнтований підхід і покладаються на людей-експертів, які працюють у певних галузях для прийняття рішень. Цей підхід використовують такі фактчекінгові сайти, як Snopes, PoltiFact, GossipCop. Ці методи є надійними, але вони вимагають багато часу і погано масштабуються з огляду на величезний обсяг контенту, доступного в соціальних мережах. Багато дослідників використовують ці сайти для створення власних наборів даних, серед них - еталонні набори даних LIAR та FakeNewsNet.

Для краудсорсингових [8] підходів «мудрість натовпу» допомагає перевірити точність новинних статей. Подібний підхід використовує Fiskkit, який надає людям платформу для обговорення важливих новинних статей і з'ясування їхньої достовірності. Хоча крауд-чекінг є відносно складним в управлінні, упередженим, з суперечливими анотаціями, менш достовірним, але має кращу масштабованість порівняно з експертним. CREDBANK - це загальнодоступний великомасштабний еталонний набір даних фейкових новин, який коментується

фактчекерами. Набори даних, створені за допомогою цього підходу, необхідно фільтрувати від користувачів, які не заслуговують на довіру, а суперечливі анотації необхідно попередньо вирішувати. Деякі подібні набори даних можна також створювати, а потім анотувати за допомогою краудсорсингових ринків, таких як AMT (Amazon Mechanical Turk).

Автоматична перевірка фактів [7], застосовуються для обробки великого обсягу даних, зокрема в соціальних мережах. Ручні підходи до перевірки фактів погано масштабуються з величезним обсягом даних, особливо згенерованих за допомогою соціальних мереж, тому для вирішення цієї проблеми були розгорнуті автоматичні методи перевірки фактів. Замість того, щоб покладатися на людський інтелект, ці методи значною мірою покладаються на обробку природної мови (NLP), інтелектуальний аналіз даних (Data Mining), машинне навчання (ML) та теорію мереж/графів.

Він використовує відкриті веб-джерела та базу знань/графік, щоб перевірити, чи є дане твердження правдивим/неправдивим. Реальні набори даних для виявлення фейкових новин зазвичай є неповними, неструктурованими, немаркованими і зашумленими, що робить автоматичне виявлення дещо складним.

Лінгвістичні шаблони та текстові ознаки [9] відрізняють фейкові новини від справжніх, зокрема через ознаки акаунтів у мережі. Існує також модель Social Article Fusion (SAF), яка використовує ознаки соціальної активності разом із лінгвістичними. Разом з цим використовуються лінгвістичні ознаки разом із синтаксичними та семантичними ознаками, щоб відрізнити справжній контент від фейкових новин. Дослідники запропонували модель для виявлення фейкових новин, яка здатна працювати з новинними твердженнями різної довжини, використовуючи різні варіації словникових вбудовувань. [1, 2]. Важливу роль має заданий новинний контент на рівні лексики, синтаксису, семантики та дискурсу. Також вивчається ієрархічна структуру документа шляхом синтаксичного аналізу залежностей на рівні речень. Не зважаючи на успіх цього методу в різних сценаріях, він створює обмеження у випадку виявлення дезінформації на популярних платформах соціальних мереж, де повідомлення короткі, а отже, лінгвістичні ознаки, витягнуті з них, часто є недостатніми для алгоритмів машинного навчання, щоб зробити точний прогноз. Крім того, ці підходи не можна використовувати для виявлення фейкових новин, які не містять текстового контенту, а лише фото або відео.

Візуальна основа [10] використовує візуальний контент, який часто розглядається як доказ, який може підвищити довіру до новинної статті, і тому видавці фейкових новин схильні використовувати провокаційний візуальний контент, щоб залучити і ввести в оману читачів. Класифіковано різні візуальні та статистичні характеристики зображень для автентифікації новин.

На основі соціального контексту [5] існує три основні аспекти соціального контексту, а саме: профілі користувачів, пости та відповіді користувачів і мережеві структури. Соціальний контекст показує, як новини поширюються в часі, і надає корисну інформацію для визначення правдивість і позицію новинних статей.

На основі мережі [4] пропонується метод виявлення фейкових новин який вивчає різні соціальні мережі, такі як мережі дружби, твітів, ретвітів, постів і репостів, щоб виявити фейкові новини. Він виявляє, хто поширює фейкові новини, зв'язки між розповсюджувачами і як фейкові новини поширюються в соціальних мережах.

Користувачі схильні створювати різні мережі на онлайн-платформах ЗМІ за спільними інтересами та схожістю, ці мережі слугують шляхами для поширення інформації. Досліджують різні мережі в соціальних мережах, що дає цінну інформацію про розповсюджувачів новин і про те, як вони пов'язані один з одним. В дослідженні [4], моделюють схему поширення повідомлень у вигляді дерева, яке разом із взаємозв'язком між публікаціями дає додаткову інформацію про часову поведінку та сентименти публікацій.

Темпоральна поведінка [11] показує, що новини в Інтернеті не є статичними, а постійно розвиваються в часі, додаючи нову інформацію або перекручуючи фактичне твердження. Це особливо помітно у випадках, коли чутки з'являються багато разів після того, як була опублікована оригінальна новина. Аналіз життєвого циклу чуток допомагає зрозуміти це явище, і тому розглядаються повторювані чутки на рівні повідомлень протягом різних періодів часу.

На основі достовірності [6] оцінюється якість новин та достовірність/переконливість, тобто визначає користувачів, які поширюють чутки, використовуючи концепцію правдоподібності. Також можливе зосередження на оцінці достовірності певного твердження. Запропоновано систему аналізу достовірності для оцінки достовірності твіту та запобігає поширенню фейкової або зловмисної інформації. TweetCred - це веб-система, яка оцінює достовірність твіту в режимі реального часу.

До наборів даних для виявлення фейкових новин висуваються такі вимоги: однорідність за довжиною, жанрами новин, темами тощо, які необхідні для створення надійного набору даних для виявлення фейкових новин, а також збір як справжніх, так і фейкових новинних статей для перевірки істинності кожного елемента набору даних.

Отже, фейкова чи неправдива інформація супроводжує життя людини скрізь в усіх сферах життєдіяльності, а тому розробка та удосконалення методів та засобів, за допомогою яких можна здійснити їх пошук, виявлення та знешкодження є актуальною задачею. Метою роботи є удосконалення методу виявлення неправдивих новин, шляхом введення класифікаторів, зокрема випадкового лісу. Це дозволить не лише підвищити точність моделі, але й забезпечити більшу стабільність та надійність результатів. Випадковий ліс є ансамблевим методом наглядуваного навчання, який використовується для задач класифікації та регресії. Він складається з набору рішень дерев, кожне з яких незалежно навчається на підмножині даних і робить прогноз. Фінальний результат моделі базується на більшості голосів (для класифікації) або середньому (для регресії) рішень всіх дерев у лісі. Також удосконалити метод проєктування модулів для знаходження та

відображення неправдивих новин потенційним використанням вебдодатку, зокрема випадковий ліс.

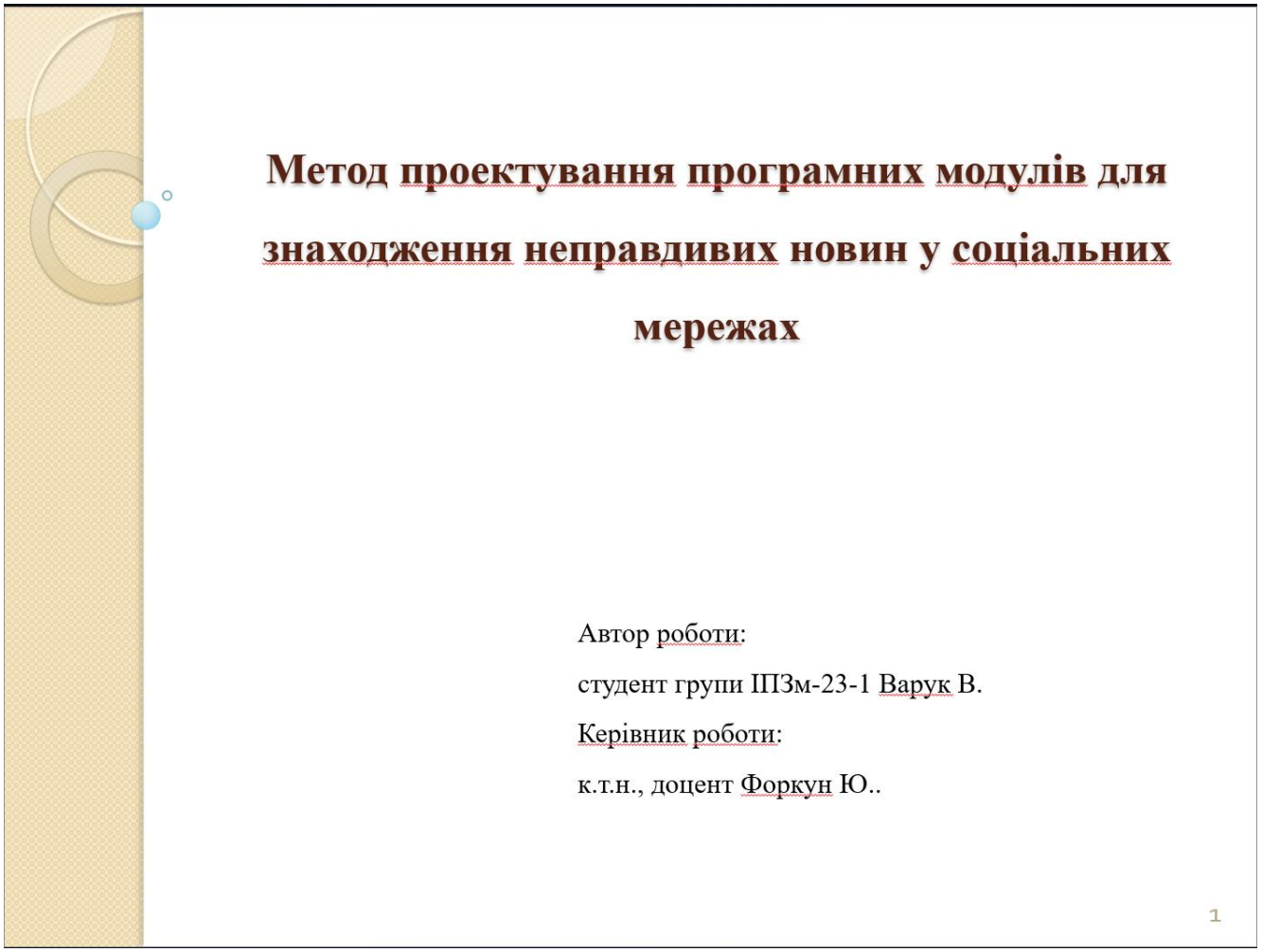
**Перелік посилань**

1. Analysis of training methods and neural network tools for fake news detection. June 2023. Cybersecurity Education Science Technique. 4(20):pp. 20-34.
2. M. H. Goldani, S. Momtazi, and R. Safabakhsh. Detecting fake news with capsule neural networks, Appl. Soft Comput., vol. 101, p. 106991, 2021.
3. Wang J, Zhu Z, Liu C, Li R, Wu X (2024) LLM-Enhanced multimodal detection of fake news. PLoS ONE 19(10): e0312240. DOI: 10.1371/journal.pone.0312240
4. Soga K, Yoshida S, Muneyasu M (2024) Graph-Based Interpretability for Fake News Detection through Topic- and Propagation-Aware Visualization. Computation 12(4): p. 82. DOI: 10.3390/computation12040082
5. Zhang C, Gupta A, Kauten C, Deokar A.V, Qin X (2019) Detecting fake news for reducing misinformation risks using analytics approaches. European Journal of Operational Research, 279(3): pp. 1036–1052. DOI: 10.1016/j.ejor.2019.02.030
6. Zhou X, Zafarani R (2020) A Survey of Fake News: Fundamental Theories, Detection Methods, and Challenges. ACM Computing Surveys 53(5): 1-40. DOI: 10.1145/3395046
7. Jwa H, Oh D, Park K, Kang J, Lim H (2019) ExBAKE: Automatic fake news detection model based on bidirectional encoder representations from transformers (BERT). Applied Sciences, 9(19): 4062. DOI: 10.3390/app9194062
8. Bondielli A, Marcelloni F (2019) A survey on fake news and rumour detection techniques. Information Sciences, 497: 38–55. DOI: 10.1016/j.ins.2019.05.035
9. Ahmed H, Traore I, Saad S (2018) Detecting Opinion Spams and Fake News Using Text Classification. Security and Privacy 1(1): e9. DOI: 10.1002/spy2.9
10. Kaliyar R.K., Goswami A., Narang P. (2020) DeepFakE: Improving fake news detection using tensor decomposition-based deep neural network. Journal of Supercomputing, 77(3): pp. 1015–1037. DOI: 10.1007/s11227-020-03284-3
11. Huang Y.F., Chen P.H. (2020) Fake news detection using an ensemble learning model based on Self-Adaptive Harmony Search algorithms. Expert Systems with Applications, 159, pp. 113-120. DOI: 10.1016/j.eswa.2020.113120

**ДОДАТОК Б**

(обов’язковий)

**ПРЕЗЕНТАЦІЙНІ МАТЕРІАЛИ**



# Актуальність теми

Актуальність теми полягає в тому, що сучасні інформаційні системи недостатньо ефективно справляються з виявленням неправдивих новин у соціальних мережах, що створює значні ризики для суспільства. Швидке поширення фейкової інформації знижує довіру до офіційних джерел, впливає на формування громадської думки та погіршує інформаційну безпеку, особливо в умовах гібридних загроз.

Використання сучасних методів машинного навчання дозволяє створювати високотехнологічні системи для автоматичного аналізу та виявлення неправдивої інформації. Такі системи здатні оперативно обробляти великі обсяги даних, виявляти складні патерни дезінформації та забезпечувати точність результатів. Таким чином, розробка методу для виявлення фейкових новин у соціальних мережах є важливим кроком у забезпеченні інформаційної безпеки та протидії дезінформації в сучасному цифровому середовищі.

# Мета і завдання дослідження

Мета дослідження полягає в розробці методу проектування програмних модулів для виявлення неправдивих новин у соціальних мережах, який дозволяє ефективно аналізувати великі обсяги інформації, виявляти потенційні фейкові новини та знижувати їхній вплив на суспільство. Метод має забезпечити високу точність визначення неправдивої інформації, автоматизацію процесу її виявлення та підвищення швидкості обробки даних. Для досягнення мети дослідження потрібно виконати такі завдання:

- Провести аналіз існуючих методів виявлення неправдивих новин у соціальних мережах
- Дослідити потенціал застосування методів машинного навчання для аналізу текстових даних
- Розробити метод проектування програмних модулів для виявлення фейкових новин
- Провести оцінку точності та ефективності роботи розробленого методу

# Об’єкт і предмет дослідження

Об’єктом дослідження є процеси проектування програмних застосунків для виявлення неправдивої інформації

Предметом дослідження є методи проектування програмних застосунків у сфері кібербезпеки та кібергігієни

# Підходи та інструменти дослідження

| Лінгвістичний аналіз тексту | Мультимодальний аналіз |
|---|---|
| Лінгвістичний аналіз тексту є основою для виявлення семантичних, граматичних та стилістичних особливостей текстового контенту, що дозволяє оцінити його достовірність. У межах дослідження були застосовані підходи, такі як морфологічний аналіз, синтаксичний аналіз, семантичний аналіз, стилістичний аналіз. | Мультимодальний аналіз дозволяє враховувати кілька видів інформації для підвищення точності визначення достовірності контенту: аналіз тексту, аналіз зображень, соціальні сигнали. |

5

# Підходи та інструменти дослідження

| Методи машинного навчання | Створення та тестування наборів даних для аналізу фейкових новин |
|---|---|
| Для автоматизації процесу виявлення фейкових новин було використано сучасні алгоритми машинного навчання, такі як BERT, SVM, Random Forest. | Одним із важливих етапів дослідження було створення спеціалізованих даних, такі як підготовка даних, розробка набору даних WarHub, тестування наборів даних. |

# Наукова новизна

Удосконалення методу проектування програмних модулів шляхом введення нових класифікаторів і проектування модулів для ефективного виявлення неправдивої інформації.

Розроблено нову модель на основі BERT, яка враховує як текстові, так і соціально-контекстуальні характеристики для аналізу достовірності інформації.

# Аналіз предметної області та існуючих рішень

Виявлення неправдивих новин у соціальних мережах є важливим завданням через значний обсяг інформації, швидкість її поширення та складність аналізу різних форматів контенту. Існуючі методи часто обмежуються текстовим аналізом, не враховуючи соціальний контекст і поведінкові патерни, що знижує їх ефективність у виявленні маніпуляцій і фейкових новин.

Недоліки традиційних підходів, таких як лінгвістичний аналіз або машинне навчання без урахування мультимодальних даних, включають низьку точність у складних сценаріях. Проблеми, зокрема використання ботів і автоматизованих систем генерації контенту, підкреслюють необхідність удосконалення методів для адаптації до сучасних викликів. Це створює потребу в розробці інноваційних рішень, які забезпечують інтеграцію мультимодального аналізу, застосування передових алгоритмів машинного навчання та врахування динаміки соціальних мереж.

# Аналіз існуючих рішень

| Підхід | Переваги | Недоліки |
|---|---|---|
| Лінгвістичні методи | Простота реалізації | Ігнорування соціального контексту |
| Методи на основі знань | Висока точність при наявності баз знань | Потреба в актуалізації баз знань |
| Соціально-контекстуальні | Виявлення шаблонів поширення фейків | Етичні питання доступу до даних |
| Гібридні методи | Комплексний аналіз | Складність реалізації |

# Концептуальний опис наборів даних

Набір даних WarHub створено для аналізу неправдивої інформації в соціальних мережах, класифікованої на категорії "фейк" та "правда". Він охоплює різноманітні теми, забезпечуючи збалансованість і універсальність, що робить його ефективним для навчання моделей.

Попередня обробка текстів включала нормалізацію, видалення стоп-слів, URL-адрес та іншого шуму, що підвищило якість даних для подальшого аналізу. Для векторизації застосовано TF-IDF, який визначає значущість слів, та Count Vectorizer, що формує вектор на основі частоти повторень. Ці заходи забезпечили готовність набору для застосування в алгоритмах машинного навчання, сприяючи ефективному виявленню дезінформації.

# Метод та алгоритм процесу визначення достовірності даних

Для визначення достовірності даних використано сучасний підхід, що поєднує попередню обробку тексту, машинне навчання та аналіз поведінкових патернів користувачів. На етапі підготовки даних текст нормалізується шляхом усунення регістрів, зайвих символів та пробілів, а також видаляється шум, зокрема стоп-слова, URL-адреси та хештеги.

Методи векторизації, такі як TF-IDF, дозволяють визначити вагу слів у контексті документу, а Count Vectorizer забезпечує представлення тексту у вигляді векторів на основі частоти слів. Ці векторизовані дані передаються в моделі машинного навчання, включаючи Random Forest, SVM та BERT, які забезпечують високоточну класифікацію на категорії "фейк" та "правда".

Алгоритми також враховують поведінкові патерни користувачів, дозволяючи виявляти аномалії у поширенні інформації. Такий підхід забезпечує адаптивність системи до змін контексту, підвищуючи її ефективність у реальних умовах.

11

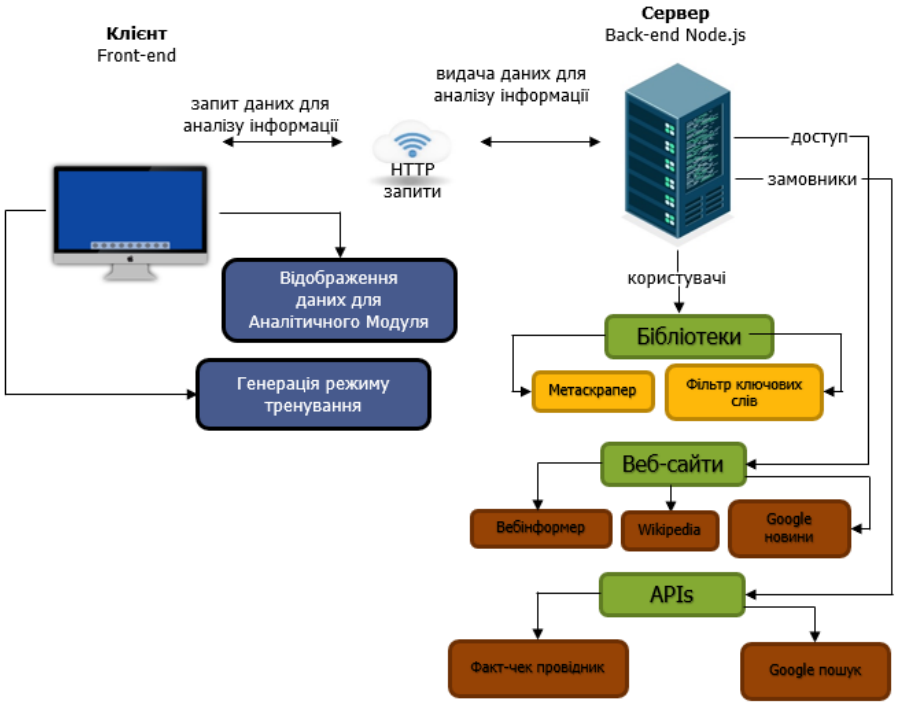
# Проектування архітектури системи

Архітектура системи розроблена з урахуванням вимог до високої продуктивності та масштабованості. Однією з ключових складових є модуль аналізу новин, який дозволяє автоматично витягувати метадані з наданих посилань, забезпечуючи швидку обробку та попередній аналіз контенту. Навчальний модуль впроваджує методи теорії щеплення, спрямовані на підвищення медіаграмотності користувачів через навчання розпізнаванню неправдивої інформації.

Для реалізації клієнт-серверної архітектури було обрано технології Node.js та Vue.js, що забезпечують гнучкість і зручність у розробці інтерактивних елементів системи. Інтеграція бібліотек Metascraper, Google Search API та Fact Check Explorer дозволяє підвищити точність перевірки фактів шляхом звернення до перевірених джерел інформації. Система також використовує модель BERT, яка базується на 12 шарах трансформаторів для глибокого контекстного аналізу тексту. У структурі моделі застосовуються токени CLS і SEP, що дозволяють здійснювати точний аналіз змісту та взаємозв'язків у тексті, забезпечуючи високу ефективність класифікації.

12

# Архітектура аналітичного модуля системи




13

# Оцінка системи визначення достовірності даних

Розроблений аналітичний модуль демонструє високий рівень точності, що досягає 99% завдяки використанню моделі BERT для аналізу даних. Ця ефективність базується на глибокому контекстному аналізі текстів, середня довжина яких у наборі даних WarHub становить менше 20 слів, що оптимізує процес класифікації. Інтеграція з такими інструментами, як Metascraper та Google Search API, дозволяє значно скоротити час перевірки фактів, підвищуючи швидкість і точність аналізу.

Практична цінність модуля полягає в його здатності забезпечувати автоматичне виявлення дезінформації у соціальних мережах. Це сприяє створенню більш надійного інформаційного середовища, зменшенню впливу неправдивих новин на громадську думку та підвищенню рівня довіри до джерел інформації.

# Наукові публікації

Варук В.К., Форкун Ю.В. Аналіз програмних методів та засобів виявлення фейкових новин, Збірник наукових працю конференції АПКН-2024, с. 73-78

# Висновки

У межах дослідження було проведено аналіз існуючих методів виявлення неправдивих новин у соціальних мережах, що дало змогу визначити їхні основні переваги та недоліки.
Було вивчено потенціал застосування методів машинного навчання для аналізу текстових даних, зокрема моделей, які враховують контекстну природу інформації.
Розроблено метод проектування програмних модулів для автоматичного виявлення фейкових новин, що базується на застосуванні сучасних алгоритмів машинного навчання та мультимодального аналізу.
Проведено оцінку точності та ефективності розробленого методу, яка підтвердила його здатність забезпечувати високу якість і точність класифікації інформації.
Розроблена система демонструє значний потенціал для боротьби з дезінформацією, що робить її цінним інструментом для покращення якості інформаційного простору.

Дякую за увагу!

Завідувачу кафедри
інженерії програмного забезпечення
проф. Леоніду БЕДРАТЮКУ
студента групи ІПЗм-23-1
Валентин ВАРУК
Ім'я, ПРІЗВИЩЕ

**ЗАЯВА**

Прошу закріпити за мною тему кваліфікаційної роботи освітнього ступеня «магістр» за спеціальністю 121 «Інженерія програмного забезпечення»:
Метод проектування програмних модулів для знаходження неправдивих новин в соціальних мережах

(керівник кваліфікаційної роботи – Юрій ФОРКУН)
Ім'я, ПРІЗВИЩЕ

02.09.24
Дата

Підпис здобувача

Завідувачу кафедри інженерії програмного
забезпечення проф. Леоніду БЕДРАТЮКУ
здобувача вищої освіти
Валентина ВАРУКА
факультет ІТ, 2 курс, група ІПЗм-23-1

**ЗАЯВА**

З правилами чинного Положення про систему забезпечення академічної доброчесності в Хмельницькому національному університеті, згідно з яким виявлення плагіату є підставою для відмови в допуску кваліфікаційної роботи до захисту і застосування заходів дисциплінарної та академічної відповідальності, ознайомлений. Про використання програмно-технічних засобів для перевірки кваліфікаційних робіт здобувачів вищої освіти на плагіат оповіщений та надаю свою згоду на обробку й збереження університетом моєї роботи в інституційному репозитарії університету.

Також надаю університету право на передачу моєї роботи для обробки та збереження в базах даних програмно-технічних засобів (Unicheck та Anti-Plagiarism) та використання роботи для виявлення плагіату в інших роботах, які перевіряються програмно-технічними засобами та користувачами, що мають доступ до цих програмно-технічних засобів, виключно в обмежених цілях для виявлення плагіату в текстах робіт.

Робота для перевірки університетом надається в друкованому та електронному варіанті. Електронна версія моєї роботи збігається (ідентична) з друкованою.

02.09.24
дата


підпис

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

**ДЕКЛАРАЦІЯ УЧАСНИКА ОСВІТНЬОГО ПРОЦЕСУ**

**щодо дотримання академічної доброчесності**

Цією декларацією я, Варук Валентин Костянтинович,

студент ІІІ курсу спеціальності 121 – Інженерія програмного забезпечення, група ІПЗс-20-1

здобувач вищої освіти (шифр та назва спец-ті, курс, академічна група)

підтверджую, що ознайомився (-лась) з Положенням про систему забезпечення академічної доброчесності у Хмельницькому національному університеті та Кодексом академічної доброчесності і **зобов'язуюсь** дотримуватися їх вимог під час освітнього процесу, проведення наукової діяльності, виконання організаційно-адміністративних функцій тощо.

**Усвідомлюю,** що у разі порушення мною принципів академічної доброчесності нестиму відповідальність перед академічною спільнотою ХНУ згідно з нормами, визначеними Положенням про систему забезпечення академічної доброчесності у Хмельницькому національному університеті, законодавства України.

05 лютого 2023 р.

Підпис

Sun Dec 01 21:04:32 EET 2024, Форкун Юрій Вікторович, Хмельницький національний університет, ХНУ

# Anti-Plagiarism v-15.257

**Максимальне співпадіння з одним документом 1.0%**

Словники перевірки: en_US, ru_RU, ua_UA. **Помилок в документах: 8%**

| ID: 152824<br>Назва: МКР_Метод проектування програмних модулів для знаходження неправдивих новин у соціальних мережах<br>Додано в БД: 2024-12-01<br>Автора: Валентин ВАРУК<br>Керівники: к. техн. наук, доцент Юрій Форкун<br>Консультанти:<br>Опоненти: | Документ | | Сумарний збіг по Базі Даних | |
|---|---|---|---|---|
| | Символи | Лексеми | Символи | Лексеми |
| | 104298 | 1558 | 1622 (2%) | 21 (1%) |

Джерело плагіату

| ID | Опис | Наявність плагіату в документі | |
|---|---|---|---|
| | | Символи | Лексеми |

**Протокол аналізу звіту подібності науковим керівником**

Заявляю, що я ознайомився (-лась) з Повним звітом подібності, який був згенерований Системою виявлення і запобігання плагіату щодо роботи:

**Автор:** Валентин ВАРУК

**Співавтор:**

**Назва:** Метод проектування програмних модулів для знаходження неправдивих новин у соціальних мережах

**Науковий керівник:** кандидат. техн. наук, доцент Нормоконтролер к. пед. наук, доцент Юрій Форкун

**Підрозділ:** Кафедра інженерії програмного забезпечення

**Коефіцієнт подібності 1:**6.1%

**Коефіцієнт подібності 2:**2.4%

**Мікропробіли:** 0

**Заміна букв:** 3

**Інтервали:** 0

**Білі знаки:** 0

**Дата створення звіту:** 2024-12-01 21:30:21.0

**Після аналізу Звіту подібності констатую наступне:**

☑ **Запозичення, виявлені в роботі є законними і не є плагіатом. Рівень подібності не перевищує допустимої межі. Таким чином робота незалежна і приймається.**

☐ **Запозичення не є плагіатом, але перевищено граничне значення рівня подібностей. Таким чином робота повертається на доопрацювання.**

☐ **Виявлено запозичення і плагіат або навмисні текстові спотворення (маніпуляції), як передбачувані спроби укриття плагіату, які роблять роботу невідповідною вимогам законодавства (Ст. 32. ЗУ Про вищу освіту, пункт 3.1, Ст. 42. ЗУ Про освіту) та вимог НАЗЯВО (Критерій 5), а також кодексу етики і процедурам. Таким чином робота не приймається.**

Обгрунтування:

2.12.2024
*Дата*

(Форкун Ю.В.)
експерт

**РІШЕННЯ** ЕКСПЕРНОЇ КОМІСІЇ
КАФЕДРИ ІНЖЕНЕРІЇ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ
ПРО ДОПУСК КВАЛІФІКАЦІЙНОЇ РОБОТИ ДО ЗАХИСТУ

Підтверджуємо ознайомлення з результатами звіту/звітів подібності щодо роботи, продукованими програмно-технічним засобом(ами) перевірки текстів на плагіат.
Назва: «Метод проектування програмних модулів для знаходження неправдивих новин у соціальних мережах»
Автор: Варук Валентин Костянтинович
Спеціальність: 121 – Інженерія програмного забезпечення
Освітня програма: Освітньо-професійна програма «Інженерія програмного забезпечення»
Науковий керівник: Форкун Юрій Вікторович, канд. техн. наук, доцент
Після аналізу звіту подібності зроблено такий висновок:

| № | Висновок | Позначка про відповідність |
|---|---|---|
| 1 | Запозичення, виявлені в роботі, є законними і не є плагіатом. Робота приймається до захисту. | **відповідає** |
| 2 | Виявлені запозичення не є плагіатом, розміщені у розділах, які не описують безпосередньо авторське дослідження, але кількість цитат перевищує обсяг, виправданий поставленою метою роботи. Робота приймається до захисту, але має бути відкоригована. Відкоригований варіант має бути поданий на кафедру за два дні до захисту, разом із заявою щодо самостійності виконання письмової роботи та ідентичності друкованої й електронної версії роботи. | |
| 3 | Виявлені запозичення не є плагіатом, але частково розміщені у розділах, які описують безпосередньо авторське дослідження, а кількість цитат перевищує обсяг, виправданий поставленою метою роботи. В зв'язку з цим мета роботи та поставлені завдання не були досягнені. Робота може бути допущена до захисту (наступного року) після того, як буде відкоригована та доопрацьована і успішно пройде повторну перевірку на академічний плагіат. | |
| 4 | Робота містить навмисні текстові спотворення, передбачувані спроби укриття запозичень або інші прояви академічного плагіату. Робота містить фабрикацію або фальсифікацію даних. Робота не допускається до захисту. | |
| 5 | Інше: | |

Підтвердження:
Запозичення, виявлені у роботі, є законними і не є плагіатом, оскільки:
1) у тексті кваліфікаційної роботи системою перевірки на плагіат StrikePlagiarism виявлено схожість з деякими документами у частині загальновживаних обов'язкових словосполучень у стандартних бланках (титулка, бланк завдання), у структурі змісту, у назвах розділів/підрозділів, у назвах публікацій переліку джерел посилання тощо;
2) в якості запозичень системою StrikePlagiarism було зафіксовано деякі послідовності вихідного коду і посиланання на бібліотеки, які є стандартними мовними конструкціями програмування та не можуть розглядатися як об'єкт авторських прав і, відповідно, їх порушення;
3) запозичення, виявлені в тексті роботи, є фрагментарними або мають належним чином оформленні посилання;
4) виявлені модифікації тексту не впливають на відсоток схожості.
Максимальний обсяг запозичень, визначений системою Anti-Plagiarism, складає 1%. За системою StrikePlagiarism коефіцієнт подібності 1 складає 6.1% коефіцеєнт подібності 2 складає 2.4%.

Дата 2.12.24

Завідувач кафедри ІПЗ _____ Леонід БЕДРАТЮК
Гарант освітньої програми _____ Оксана ЯШИНА
Керівник кваліфікаційної роботи _____ Юрій ФОРКУН

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

**РЕЦЕНЗІЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ**
освітнього ступеня «Магістр»

Дипломник Варук Валентин Костянтинович

Тема Метод проектування програмних модулів для знаходження неправдивих новин у соціальних мережах.

Спеціальність 121 – Інженерія програмного забезпечення

**Обсяг кваліфікаційної роботи:**

Кількість листів креслень 0; кількість сторінок записки 106

1.Короткий зміст пояснювальної записки та прийнятих рішень У кваліфікаційній роботі було досліджено і проаналізовано предметну область. Проведено детальний аналіз існуючих рішень та практик, що стосуються виявлення неправдивої інформації, визначено їх переваги та недоліки. У рамках роботи було сформовано технічне завдання, визначено необхідні технології для удосконалення методу, проведено опис архітектурного та структурованого рішення, а також проектування модуля системи виявлення неправдивої інформації. Здійснено удосконалення методу проектування програмних модулів шляхом введення нових класифікаторів і проектування модулів для ефективного виявлення неправдивої інформації. Також було реалізовано запропонований метод у вигляді програмного рішення та проведено його тестування для перевірки функціональності та ефективності.

2. Висновок про відповідність проекту поставленому завданню Кваліфікаційна робота виконана відповідно до поставленого завдання та з дотриманням всіх вимог.

3. Характеристика виконання кожного розділу проекту, ступінь використання останніх досягнень науки і техніки та передових методів роботи У вступі доведено актуальність теми, визначено мету та завдання кваліфікаційної роботи. У першому розділі проведено аналіз предметної області, а також реальний та актуальний стан проблеми, що досліджується, проаналізовано метод рішення проблеми. У другому розділі здійснено удосконалення методу виявлення неправдивих новин. У третьому розділі було здійснено опис архітектурного та структурного рішення, а також проектування модуля системи виявлення неправдивої інформації. У четвертому розділі здійснено теоретичне обґрунтування та описано можливу теоретичну та практичну перевірку потенційними користувачами. У цьому розділі представлені як методичні рішення, так і результати цього процесу.

4. Позитивні сторони проекту Тематика кваліфікаційної роботи є актуальною, оскільки новини завжди були критичною і складною проблемою в інформаційному середовищі. Поширення неправдивих новин викликає серйозне занепокоєння, особливо у сфері медичної інформації, що може мати небезпечні та потенційно смертельні наслідки. Також було застосовано новітні технології для удосконалення методу та актуальні архітектурні рішення.

5. Негативні сторони проекту Відсутні

6. Оцінка графічного оформлення та пояснювальної записки проекту Графічне оформлення виконано відповідно до теми кваліфікаційної роботи та подано у вигляді діаграм і рисунків. Пояснювальна записка оформлена згідно вимог чинних стандартів.

7. Відгук про кваліфікаційну роботу в цілому Кваліфікаційна робота заслуговує позитивної оцінки. Викладення матеріалу пояснювальної записки є структурованим, послідовним та чітким, що дозволяє зрозуміти викладений матеріал у рамках тематики кваліфікаційного проекту. Графічний матеріал надає можливість наочно побачити деталі проектування методу.

8. Інші зауваження

9. Оцінка кваліфікаційної роботи Кваліфікаційна робота виконана у повному обсязі, відповідає поставленій задачі та заслуговує на оцінку «добре», С.

РЕЦЕНЗЕНТ к.т.н., доцент кафедри КІІС, Капустян Марія Вікторівна

"02" грудня 2024 р. _____
(підпис)