МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

**ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ**

**НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ЗАХИСТУ ІНФОРМАЦІЇ
КАФЕДРА УПРАВЛІННЯ ІНФОРМАЦІЙНОЮ ТА КІБЕРНЕТИЧНОЮ БЕЗПЕКОЮ**

КВАЛІФІКАЦІЙНА РОБОТА

на тему: “РОЗРОБЛЕННЯ МЕТОДИКИ ОЦІНЮВАННЯ ЗАХИСТУ ВІД РОЗПОВСЮДЖЕННЯ ФЕЙКОВИХ НОВИН ТА ДЕЗІНФОРМАЦІЇ В СОЦІАЛЬНИХ МЕРЕЖАХ”

на здобуття освітнього ступеня бакалавра
зі спеціальності 125 Кібербезпека
освітньої програми Управління інформаційною та кібернетичною безпекою

Кваліфікаційна робота містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело

_____ Мирослава НОРЕНКО
(підпис) *Ім’я, ПРІЗВИЩЕ здобувача*

Виконала: здобувачка вищої освіти гр. УБД-41

Мирослава НОРЕНКО
Ім’я, ПРІЗВИЩЕ

Керівник: Олександр ПОРОХНИЦЬКИЙ
Д.е.н., проф. Ім’я, ПРІЗВИЩЕ

Рецензент:
Д.т.н., проф. Ім’я, ПРІЗВИЩЕ

Київ 2024

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ

Навчально-науковий інститут захисту інформації

Кафедра управління інформаційною та кібернетичною безпекою

Ступінь вищої освіти бакалавр

Спеціальність 125 Кібербезпека

Освітня програма Управління інформаційною та кібернетичною безпекою

ЗАТВЕРДЖУЮ
Завідувач кафедри УІКБ
_____ Світлана ЛЕГОМІНОВА
“____”_____2024 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

Норенко Мирославі Олексіївні
(прізвище, ім’я, по батькові здобувача)

1. Тема кваліфікаційної роботи “розроблення методики оцінювання захисту від розповсюдження фейкових новин та дезінформації в соціальних мережах”,
керівник кваліфікаційної роботи ПОРОХНИЦЬКИЙ Олександр,
(ПРІЗВИЩЕ, Ім’я,, науковий ступінь, вчене звання)

затверджені наказом Державного університету інформаційно-комунікаційних технологій від 27.02.24 № 36.

2. Строк подання кваліфікаційної роботи “20” травня 2024р.

3. Вихідні дані до кваліфікаційної роботи: *міжнародні стандарти, наукова та технічна література. методи та засоби шифрування даних, загрози та вразливості безпеки хмарних обчислень*

4. Перелік питань, які мають бути розроблені:
4.1. Проаналізувати методи оцінювання захисту від дезінформації та фейкових новин.
4.2. Розробити методики оцінювання захисту від розповсюдження фейкових новин та дезінформації в соціальних мережах
4.3. Практичне застосування розробленої методики

5. Перелік ілюстративного матеріалу: *презентація PowerPoint*

6. Дата видачі завдання “11“ березня 2024 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Етапи кваліфікаційної роботи	Термін виконання етапів роботи	Примітка
1.	Визначення об’єкту, предмету, мети та завдань дослідження.	18.03.2024	
2.	Збір та аналіз літератури.	29.03.2024	
3.	Аналіз дезінформації та фейкових новини, методів оцінювання захисту.	08.04.2024	
4.	Розроблення методики оцінювання захисту.	22.04.2024	
5.	Практичне застосування розробленої методики.	08.05.2024	
6.	Формулювання висновків за результатами проведеного дослідження.	20.05.2024	
7.	Оформлення роботи.	22.05.2024	
8.	Оформлення презентації.	03.06.2024	
9.	Отримання рецензії на роботу.	03.06.2024	
10.	Захист в ДЕК.	__.06.2024	

Здобувачка вищої освіти _____ Мирослава НОРЕНКО
(підпис) *(Ім’я, ПРІЗВИЩЕ)*

Керівник кваліфікаційної роботи _____ Олександр ПОРОХНИЦЬКИЙ
(підпис) *(Ім’я, ПРІЗВИЩЕ)*

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ
НАВЧАЛЬНО-НАУКОВИЙ ІНСТИТУТ ЗАХИСТУ ІНФОРМАЦІЇ

ПОДАННЯ
ГОЛОВІ ЕКЗАМЕНАЦІЙНОЇ КОМІСІЇ
ЩОДО ЗАХИСТУ КВАЛІФІКАЦІЙНОЇ РОБОТИ
на здобуття освітнього ступеня бакалавра

Направляється здобувач Норенко М.О. до захисту кваліфікаційної роботи
(прізвище та ініціали)
за спеціальністю 125 Кібербезпека
(код, найменування спеціальності)
освітньої програми Управління інформаційною та кібернетичною безпекою
(назва)
на тему: "Розроблення методики оцінювання захисту від розповсюдження фейкових новин та дезінформації в соціальних мережах"
Кваліфікаційна робота і рецензія додаються.

Директор ННІЗІ _____ Віталій САВЧЕНКО
(підпис) *(Ім'я, ПРІЗВИЩЕ)*

Висновок керівника кваліфікаційної роботи

ВІДГУК РЕЦЕНЗЕНТА
на кваліфікаційну бакалаврську роботу

РЕФЕРАТ

Кваліфікаційна робота присвячена аналізу та розробленню методів оцінювання захисту від розповсюдження фейкових новин та дезінформації в соціальних мережах. Робота складається зі вступу, трьох розділів, що містять 16 рисунків, висновків і списку використаних джерел із 47 найменувань. Загальний обсяг роботи становить 70 аркушів, з яких 6 аркушів займають перелік умовних скорочень та список використаних джерел.

Метою роботи є розробка та апробація методики оцінювання захисту від розповсюдження фейкових новин у сучасному інформаційному просторі. Це включає створення ефективного інструментарію для виявлення та аналізу фейкових новин, а також оцінку рівня захищеності інформаційних ресурсів.

Об'єктом дослідження є сучасний інформаційний простір, зокрема соціальні мережі, новинні портали та інші онлайн-платформи, які є основними каналами розповсюдження інформації. Цей простір характеризується високою динамічністю, великою кількістю користувачів та різноманітністю контенту, що створює як можливості, так і виклики для боротьби з фейковими новинами.

Предмет дослідження є методики оцінки захисту від фейкових новин, які включають критерії, інструменти та алгоритми для виявлення та аналізу фейкових новин, а також оцінку ефективності існуючих інформаційних ресурсів у контексті протидії дезінформації.

Методи дослідження. Опис методів, які будуть використовуватися в роботі

1. Аналіз літератури — проведення огляду наукової літератури, досліджень та статей, присвячених проблемі фейкових новин та методикам їх виявлення і оцінювання захисту.

2. Експериментальні методи — розробка та проведення експериментів для тестування розроблених алгоритмів та методик на реальних даних з соціальних мереж та новинних порталів.

3. Моделювання — створення математичних та комп'ютерних моделей для симуляції процесів розповсюдження фейкових новин та оцінки ефективності захисних заходів.

4. Методи збору та аналізу даних — використання інструментів для збору даних з різних онлайн-платформ, а також методів обробки та аналізу цих даних для виявлення фейкових новин.

5. Статистичний аналіз — застосування статистичних методів для оцінки результатів тестування та аналізу ефективності розробленої методики.

6. Методи машинного навчання — використання алгоритмів машинного навчання для автоматичного виявлення фейкових новин та оцінки їх впливу на інформаційне середовище.

Галузь застосування. Розроблені рекомендації можуть бути використані при плануванні та реалізації покращення методів оцінювання захисту від розповсюдження фейкових новин та дезінформації в соціальних мережах.

Ключові слова: ФЕЙКОВІ НОВИНИ, ДЕЗІНФОРМАЦІЯ, СОЦІАЛЬНІ МЕРЕЖІ, МЕТОДИ ЗАХИСТУ

ABSTRACT

The qualification work is devoted to the analysis and development of methods for assessing protection against the spread of fake news and disinformation in social networks. The work consists of an introduction, three chapters containing 16 figures, conclusions and a list of references of 47 titles. The total volume of the work is 70 pages, of which 6 pages are occupied by the list of abbreviations and the list of references.

The purpose of the work is to develop and test a methodology for assessing protection against the spread of fake news in the modern information space. This includes creating effective tools for identifying and analyzing fake news, as well as assessing the level of security of information resources.

The object of the study is the modern information space, in particular social networks, news portals and other online platforms, which are the main channels of information dissemination. This space is characterized by high dynamism, a large number of users and a variety of content, which creates both opportunities and challenges for combating fake news.

The subject of the study is methods for assessing protection against fake news, which include criteria, tools and algorithms for identifying and analyzing fake news, as well as assessing the effectiveness of existing information resources in the context of countering disinformation.

Research methods. Description of the methods to be used in the work

1. Literature analysis - a review of scientific literature, studies and articles on the problem of fake news and methods of detecting and evaluating protection.

2. Experimental methods - developing and conducting experiments to test the developed algorithms and methods on real data from social networks and news portals.

3. Modeling - creation of mathematical and computer models to simulate the processes of fake news spreading and evaluate the effectiveness of protective measures.

4. Data collection and analysis methods - the use of tools for collecting data from various online platforms, as well as methods for processing and analyzing this data to identify fake news.

5. Statistical analysis - the use of statistical methods to evaluate test results and analyze the effectiveness of the developed methodology.

6. Machine learning methods - the use of machine learning algorithms to automatically detect fake news and assess its impact on the information environment.

Scope of application. The developed recommendations can be used in planning and implementing improved methods for assessing protection against the spread of fake news and disinformation in social networks.

Keywords: FAKE NEWS, DISINFORMATION, SOCIAL NETWORKS, METHODS OF PROTECTION

ЗМІСТ

ВСТУП

Актуальність теми. Визначення проблеми фейкових новин у сучасному інформаційному просторі.

У сучасному світі інформація відіграє ключову роль у житті суспільства. Розвиток інформаційних технологій та широке використання інтернету призвели до кардинальних змін у способах отримання, передачі та споживання інформації. Водночас, поряд зі збільшенням обсягів доступної інформації, виникла проблема розповсюдження фейкових новин. Фейкові новини (fake news) – це свідомо неправдива або маніпулятивна інформація, створена з метою ввести в оману читачів або аудиторію. Їхня поява може мати серйозні наслідки для суспільства, включаючи дезінформацію громадськості, підрив довіри до засобів масової інформації, політичну нестабільність, а також вплив на економічні ринки та особисту безпеку громадян.

Фейкові новини швидко розповсюджуються через соціальні мережі та інші онлайн-платформи, що ускладнює їх виявлення та боротьбу з ними. Завдяки анонімності в інтернеті та можливості швидкого створення та поширення контенту, фейкові новини можуть набувати широкого розголосу, створюючи ілюзію достовірної інформації. У таких умовах суспільство потребує ефективних інструментів та методик для оцінки та протидії фейковим новинам.

Важливість розробки методик для оцінки захисту від фейкових новин. Розробка методик для оцінки захисту від фейкових новин є вкрай важливою в сучасному інформаційному середовищі. Вони дозволяють не лише виявляти неправдиву інформацію, але й оцінювати рівень захищеності інформаційних ресурсів від впливу дезінформації. Такі методики можуть включати різноманітні підходи, починаючи від використання алгоритмів

машинного навчання для автоматичного виявлення фейкових новин, до розробки систем перевірки фактів та аналізу джерел інформації.

Ефективна методика оцінки захисту від фейкових новин дозволяє:

- Підвищити обізнаність громадян щодо загроз, пов'язаних з фейковими новинами.
- Зменшити вплив дезінформації на суспільство.
- Підвищити рівень критичного мислення та медіаграмотності серед населення.
- Забезпечити ефективну роботу засобів масової інформації та інших інформаційних ресурсів.

Таким чином, розробка та впровадження методик для оцінки захисту від фейкових новин є актуальним завданням, яке потребує міждисциплінарного підходу та активної співпраці фахівців у галузях інформаційних технологій, соціальних наук, журналістики та інших напрямів.

Мета роботи є розробка та апробація методики оцінювання захисту від розповсюдження фейкових новин у сучасному інформаційному просторі. Це включає створення ефективного інструментарію для виявлення та аналізу фейкових новин, а також оцінку рівня захищеності інформаційних ресурсів.

Основні завдання, які потрібно вирішити для досягнення мети

1. Провести аналіз сучасних досліджень і підходів до виявлення та боротьби з фейковими новинами.
2. Визначити критерії та показники, які можуть бути використані для оцінки захисту від фейкових новин.
3. Розробити методологію збору та аналізу даних для оцінки захисту від фейкових новин.
4. Створити модель оцінювання захисту від фейкових новин, яка включатиме алгоритми та інструменти для виявлення фейкових новин.

5. Провести тестування та верифікацію розробленої методики на обраному інформаційному середовищі (соціальні мережі, новинні портали тощо).

6. Оцінити ефективність розробленої методики та надати рекомендації щодо її вдосконалення.

Об'єкт дослідження **–** є сучасний інформаційний простір, зокрема соціальні мережі, новинні портали та інші онлайн-платформи, які є основними каналами розповсюдження інформації. Цей простір характеризується високою динамічністю, великою кількістю користувачів та різноманітністю контенту, що створює як можливості, так і виклики для боротьби з фейковими новинами

Предмет дослідження – методики оцінки захисту від фейкових новин, які включають критерії, інструменти та алгоритми для виявлення та аналізу фейкових новин, а також оцінку ефективності існуючих інформаційних ресурсів у контексті протидії дезінформації.

Методи дослідження. Опис методів, які будуть використовуватися в роботі

1. Аналіз літератури — проведення огляду наукової літератури, досліджень та статей, присвячених проблемі фейкових новин та методикам їх виявлення і оцінювання захисту.

2. Експериментальні методи — розробка та проведення експериментів для тестування розроблених алгоритмів та методик на реальних даних з соціальних мереж та новинних порталів.

3. Моделювання — створення математичних та комп'ютерних моделей для симуляції процесів розповсюдження фейкових новин та оцінки ефективності захисних заходів.

4. Методи збору та аналізу даних — використання інструментів для збору даних з різних онлайн-платформ, а також методів обробки та аналізу цих даних для виявлення фейкових новин.

5. Статистичний аналіз — застосування статистичних методів для оцінки результатів тестування та аналізу ефективності розробленої методики.

6. Методи машинного навчання — використання алгоритмів машинного навчання для автоматичного виявлення фейкових новин та оцінки їх впливу на інформаційне середовище.

Практичне значення одержаних результатів. Застосування напрацювань дасть змогу оцінювати захист від розповсюдження фейкових новин та дезінформації в соціальних мережах.

Апробація результатів кваліфікаційної роботи відбулася на Всеукраїнській науково-практичній конференції “Стратегії кіберстійкості: управління ризиками та безперервність бізнесу” 28 лютого 2024 року.

РОЗДІЛ 1 ДЕЗІНФОРМАЦІЯ ТА ФЕЙКОВІ НОВИНИ: СУТНІСТЬ, МЕХАНІЗМИ РОЗПОВСЮДЖЕННЯ ТА ПІДХОДИ ДО БОРОТЬБИ

1.1 Сутність та загальні поняття про дезінформацію та фейки

Швидкий розвиток інформаційних технологій у сфері масової комунікації породжує нові інформаційні виклики та загрози як для України, так і для всього світу. Одним із найскладніших і найпідступніших сьогодні є дезінформація.

Кембриджський словник називає дезінформацію “неправдивою інформацією, яка поширюється з метою введення в оману людей” [1]. У Спільній декларації про свободу вираження думки, “фейкові новини”, дезінформацію і пропаганду представники ООН, ОБСЄ та інших організацій зазначили, що найчастіше дезінформація спрямована на обман людей і перешкоджання знати, отримувати, шукати, поширювати інформацію [2]. А Спеціальний доповідач із заохочення і захисту права на свободу думок і їхнє вільне вираження (ООН) у доповіді від 13 квітня 2021 року узагальнив поняття дезінформації як “брехливої інформації, навмисне поширюваної з метою заподіяння серйозної соціальної шкоди” [3].

Кодекс практики ЄС щодо протидії дезінформації вказує, що дезінформація має на меті економічні вигоди для поширювача [4]. Європейська Комісія у своєму зверненні до інших органів ЄС щодо протидії онлайн-дезінформації представила аналогічне бачення сутності дезінформації, під якою розуміють очевидно неправдиву або таку, яка вводить в оману, інформацію, що в сукупності створена, представлена і поширена з метою економічної вигоди або умисного введення в оману громадськості.

Європейські експерти детальніше визначили наслідки використання дезінформації на шкоду суспільству, зокрема через загрозу демократичним політичним процесам і процесам вироблення політики, а також таким суспільним благам як захист здоров'я громадян, довкілля і безпека [5]. Відповідно до концепції ЮНЕСКО [6], поняття дезінформації (dis-information) розглядається у взаємозв'язку з іншими двома категоріями: недостовірна інформація (mis-information) і шкідлива інформація (mal-information) (Рис. 1).

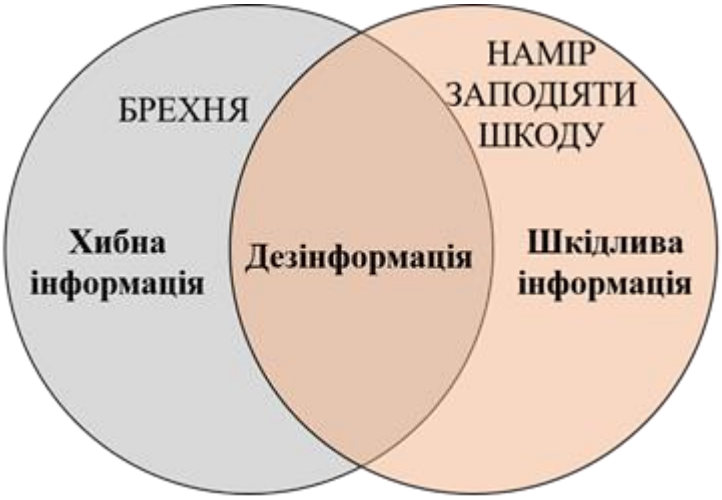


Рис. 1.1. Співвідношення понять дезінформація, хибна та шкідлива інформація [6]

Таким чином, у таблиці 1 узагальнено теоретичні приклади, які використовуються для з'ясування ознак, що містяться в поняттях хибна інформація, дезінформація та шкідлива інформація.

Таблиця 1.1

Тлумачення хибної інформації, дезінформації та шкідливої інформації

Тип	**Підходи щодо тлумачення хибна інформація, дезінформація та шкідлива інформація**
ХИБНА ІНФОРМ	"Хибна інформація є різновидом інформації, так само як і хибна інформація є різновидом інформування... інформування не вимагає правди, і інформація не

Тип	Підходи щодо тлумачення хибна інформація, дезінформація та шкідлива інформація
	обов'язково повинна бути правдивою; але хибна інформація вимагає брехні, і дезінформація повинна бути брехливою” [7].
	“неправдива інформація", тобто хибна інформація, є лише псевдоінформацією” [8].
	“хибна інформація - це "добре сформовані та осмислені дані (тобто смисловий зміст), які є неправдивими” [9].
	“Хибна інформація також може бути невизначеною (можливо, шляхом представлення більш ніж однієї можливості або вибору), розпливчастою (нечіткою) або двозначною (відкритою для різних інтерпретацій). Хибна інформація, однак, може бути правдивою, точною та інформативною, залежно від контексту, а отже, відповідати багатьом з тих самих критеріїв, які прийняті для інформації” [10].
	“Неточна інформація (або хибна інформація) може вводити людей в оману, незалежно від того, чи є вона результатом чесної помилки, недбалості, несвідомого упередження або (як у випадку дезінформації) навмисного обману” [11].
ДЕЗІНФОРМАЦІЯ	“Перш за все, для того, щоб дезінформувати, потрібно мати намір когось обдурити”. “Варто також зазначити, що потрібно мати намір обманути, а не просто мати намір поширити неправдиву інформацію” [12].
	“Дезінформація” - це просто хибна інформація, цілеспрямовано передана, щоб ввести одержувача в оману і змусити його повірити, що це інформація" [9].
	“Дезінформація виникає щоразу, коли процес інформування є дефектним. Це може статися через: (а) брак об'єктивності, як у випадку пропаганди; (б) брак повноти, як у випадку “прокляття пам'яті” (damnatio memoriae); і (в) брак плюралізму, як у випадку цензури” [13].
	“Disinformation is deliberately deceptive information. The intentions behind such deception are unknowable, but may include socially- motivated, benevolent reasons […] and personally-motivated, antagonistic reasons” [14].
	“Дезінформація - це свідомо оманлива інформація. Наміри, що стоять за таким обманом, невідомі, але можуть включати соціально мотивовані, доброзичливі причини [...] та особистісно мотивовані, антагоністичні причини” [15].
	“Дезінформація є особливо небезпечною, тому що людей не випадково вводять в оману. Дезінформація виходить від того, хто активно бере участь у спробі ввести в оману” [11].
ШКІДЛИВА ІНФОРМАЦІЯ	“правдива інформація, яка поширюється з метою заподіяння шкоди” [16].
	“потенційно небезпечна або шкідлива інформація; недоречна інформація; інформація, від якої люди відчувають дискомфорт у відкритому доступі” [17].

Тип	Підходи щодо тлумачення хибна інформація, дезінформація та шкідлива інформація
	“інформація, яка ґрунтується на реальних фактах, але використовується для завдання шкоди особі, організації або країні” [18].
	“Шкідлива інформація вимагає як наміру, так і еквівалентності, і часто включає в себе переосмислення правдивої цінності інформації в оманливих цілях” [19].

Таким чином, дезінформація — це недостовірна, оманлива, маніпулятивна інформація, створена навмисно заради отримання економічних, політичних або інших вигод. Важливо звернути увагу, що процес поширення дезінформації називається дезінформуванням. Серед форм дезінформації виділяють текстовий контент, відеоконтент, аудіальний контент, а серед методів її поширення — координована неавтентична поведінка, таргетинг, діпфейки [20], а також фейкові новини.

В Україні слово фейк звучить з екранів телевізорів, з текстів ЗМІ. Але немає чіткого його визначення і типології. Часто фейком називають недостовірну, неправдиву інформацію, неперевірений фактаж. Вважаємо, що ці поняття не відображають суті фейку. Адже фейк – це підробка, фальшивка, яка розповсюджується спеціально для того, щоб дезінформувати аудиторію.

Згідно з Barclay, фейкові новини - це "інформація, яка повністю сфабрикована з метою або заробітку, або просування певного політичного чи соціального порядку денного, як правило, шляхом дискредитації інших" [21]. Однак важливо підкреслити, що термін "фейкові новини" використовується політиками "як зброя для нападу на вільну і незалежну пресу", як зазначає Wardle [22]. У цьому сенсі Рубін, Чен і Конрой класифікують три типи фейкових новин: очевидні вигадки, містифікації та новинна сатира [23].

Прикладом очевидної фальсифікації є жовта преса та її неперевірені статті, які за допомогою клікбейтів та сенсаційних статей мають на меті

збільшити свій трафік і, як наслідок, отримати прибуток. Містифікації. Згідно з Рубіном, Ченом і Конроєм, містифікація - це "ще один тип навмисної вигадки або фальсифікації в мейнстрімі або соціальних мережах" [23]. Прикладами містифікацій є чутки, підроблені графіки чи таблиці, фальшиве приписування авторства, драматичні образи тощо. Новинна сатира або пародія. Можна знайти як гумористичні новинні веб-сайти, засновані на іронії, часто в мейнстрімному форматі. У деяких випадках, якщо читачі не знають про гумористичний ухил, такі новини можуть бути джерелом дезінформації. Важливо зазначити, що їх не слід плутати з веб-сайтами-самозванцями, які навмисно намагаються ввести в оману або заплутати, копіюючи традиційні ЗМІ. Фейкові відгуки. Щодо фейкових відгуків як інструменту дезінформації, то приклади можна знайти на платформах електронної комерції [24], де їх використовують для впливу на купівлю товарів і послуг. У зв'язку з цим автори продемонстрували, що люди не завжди здатні розпізнати оманливі думки. Випадки фальшивих рецензій також можна знайти в процесі рецензування в науковій комунікації, як, наприклад, випадок шахрайського рецензування, який призвів до відкликання трьох статей одних і тих самих авторів [25].

1.2 Аналіз медіаспоживання в інформаційному просторі України

Початок широкомасштабної збройної агресії Російської Федерації проти України призвів до різкого зростання використання соціальних мереж як джерела новин, а також посилення ролі традиційних засобів масової інформації. Українське суспільство фактично живе в інформаційному просторі та задовольняє свої потреби новинним контентом. Інформація про міжнародні події, зовнішню та внутрішню політику нашої держави, повітряні

тривоги, зміни в оперативній обстановці на фронті – все це стало звичною частиною життя багатьох українців.

Однак за останні роки змінилися не лише звички споживання інформації, як інформаційного продукту, але й сам інформаційний простір зазнав значних трансформацій. Наразі Україна перебуває в активній фазі інформаційної війни, яка проявляється в поширенні інформаційно-психологічних спецоперацій, дезінформаційних кампаній та використанні автоматизованих акаунтів для впливу на громадську думку. Український уряд докладає зусиль для швидкого реагування на ці виклики, але водночас мимоволі сприяє монополізації інформаційного простору, що призводить до браку довіри в українському суспільстві.

Саме тому критичним залишається розуміння специфіки медіаспоживання різних цільових аудиторій українського суспільства. Водночас будь-яка боротьба з дезінформацією, поширення достовірних новин та офіційних звернень будуть неефективними, якщо відбуватимуться не на тих платформах, де суспільство справді споживає інформацію.

Аналізуючи споживання новин у 2023 році, можна зазначити, що серед майже всіх видів медіа воно залишилося на рівні минулих років, водночас використання телебачення продовжує зменшуватися. Майже 47% користувачів використовують для отримання новин кілька джерел. Ті, хто використовують лише одне джерело, як правило, віддають перевагу соціальним мережам.

Слід відмітити тенденцію, яка вказує на те, що кількість людей, які щодня користуються інтернетом, неухильно зростає з кожним роком. Так, у 2023 році цей показник сягнув 89%, а серед користувачів віком від 18 до 35 років – 98%. Споживачі новин в інтернеті (інтернет-медіа та соціальних

мережах), які не дивляться новини на ТБ, складають основну частину цільової аудиторії, яка становить 66% [26].

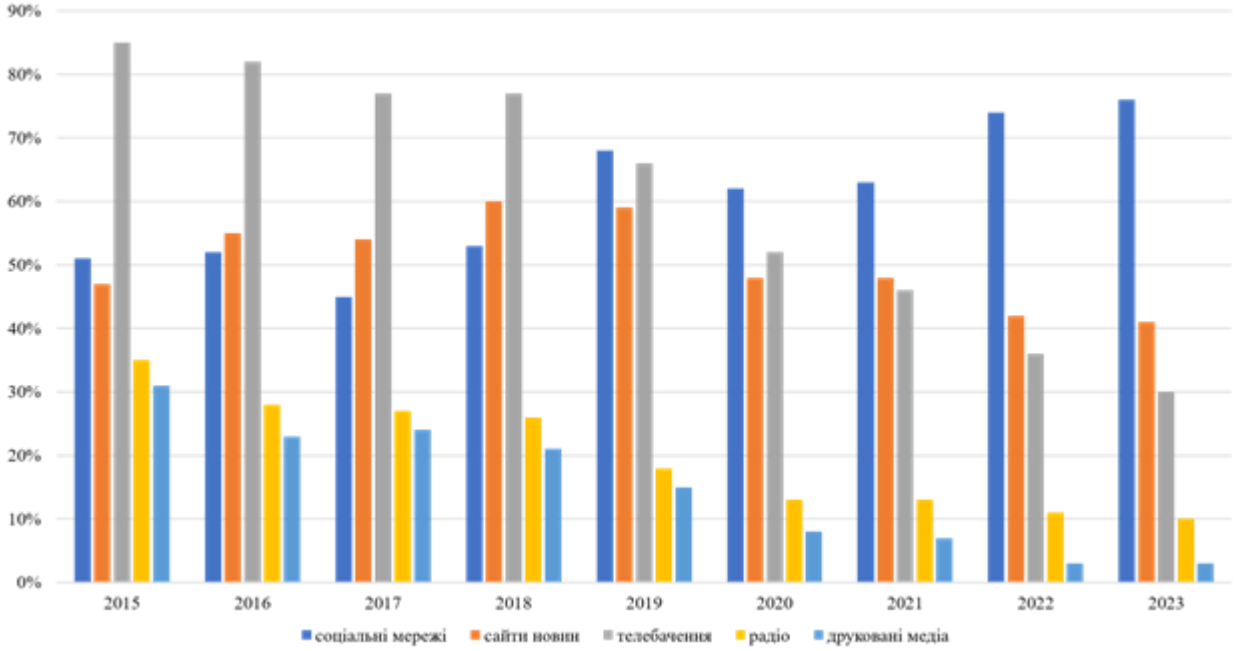


Рис. 1.2. Використання медіа для отримання новин протягом 2015-2023 років [26]

Основним девайсом для отримання інформації є смартфон, а головним джерелом новин – Telegram-канали. Окрім того, використовуються інші соціальні мережі, різні джерела в YouTube (телевізійні канали, які транслюються через YouTube, персональні канали лідерів думок/експертів), інтернет-сайти новин. Якщо порівнювати рейтинг найпопулярніших соціальних мереж в нашій державі з минулого року, то найбільше українці використовують Telegram (71,3%), YouTube (66,2%) та Facebook (55%). Також 50% опитаних споживають новини у Viber, 29,5% — в Instagram, 25,1% — у TikTok, 8,3% — у Twitter [27].

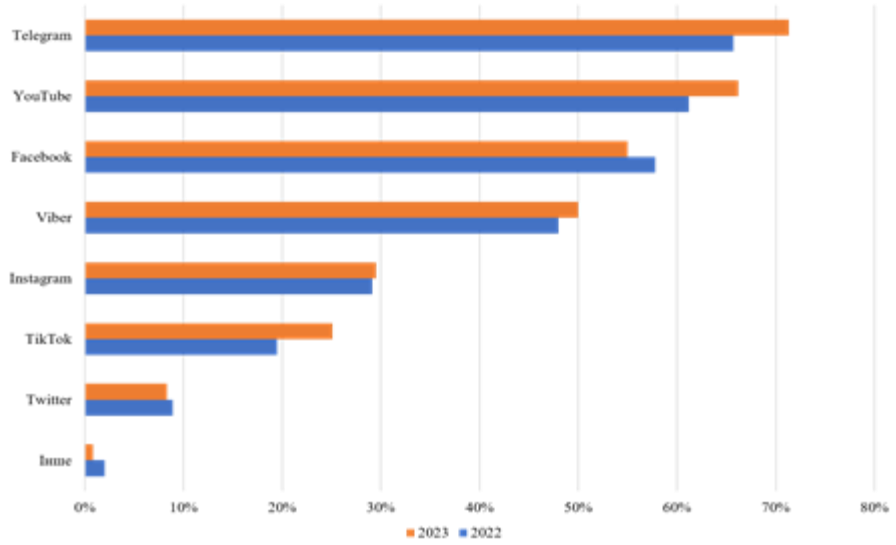


Рис. 1.3. Рейтинг найпопулярніших соціальних мереж [27]

Втім, соціологічне опитування показало, що різні вікові категорії обирають різні соціальні мережі. Так, найпопулярнішими серед молоді у віці 18-29 років залишаються Telegram (89,9%), YouTube (64,5%) та Instagram (46,1%), а от Facebook (34,1%) та Viber (27,7%) за звітний період втратили популярність. Натомість користувачі віком 30-39 років надають перевагу Telegram (75,6%), YouTube (61,1%) та Facebook (59,1%), хоча останній за минулий рік втратив майже 5% своєї аудиторії такого віку. Серед 40-49-річних Telegram (70,9%), YouTube (65,6%) та Facebook (64%) також упевнено тримають перевагу, але тут трійку лідерів наздоганяє Viber із показником у 54,2%. Користувачі віком 50-59 років для отримання новин найчастіше використовують Telegram (70%), YouTube (69,3%), Facebook (59,6%) та Viber (59,5%). Водночас серед опитаних віком 60–69 років на перший план виходять YouTube (71,3%), Viber (68%) та Facebook (62,4%), а от Telegram опинився лише на четвертій сходинці із 53,5% користувачів такого віку. Серед аудиторії віком від 70 років найбільш популярні YouTube (72,7%) та Viber (58,3%), а рівень використання інших соцмереж для споживання новин не перевищує

40%. Показово, що TikTok, хоч і не отримав переваги серед жодної з вікових категорій, демонструє найбільше зростання частки прихильників. За рік його аудиторія віком 70+ зросла з 12,2% до 21,1%, а серед людей віком 40-49 років — із 17,9% до 26,5%.

Рис. 1.4. Пристрої для отримання новин [24]

Таким чином, соціальні мережі швидко набувають популярності в Україні. Щороку кількість користувачів збільшується, охоплюючи значну і різносторонню цільову аудиторію. Так, соціальні мережі перетворюються на інструментарій ведення інформаційних операцій, здійснюючи інформаційний (психологічний) вплив на громадськість, а також відіграють безпосередню роль в інформуванні населення та можуть бути головним інструментом під час реалізації заходів стратегічних комунікацій.

1.3 Механізми розповсюдження фейкових новин та дезінформації

Фейкові новини та дезінформація розповсюджуються через різні канали та використовують різні механізми для досягнення максимальної аудиторії. Основні механізми включають:

Соціальні мережі - основний канал розповсюдження фейкових новин та дезінформації завдяки їхній популярності та можливості швидкого поширення контенту серед великої кількості користувачів [28]. Наприклад, під час виборів в США у 2016 році фейкові новини активно поширювалися через Facebook, створюючи враження масової підтримки певних політичних сил.

Інтернет-форуми та блоги – платформи, де користувачі можуть ділитися інформацією без належної перевірки фактів. Наприклад, на форумах Reddit часто з'являються неперевірені новини, які можуть викликати паніку або неправдиві висновки серед користувачів.

Месенджери – особисті повідомлення, які дозволяють швидко передавати новини у закритих групах або між окремими особами. Наприклад, під час пандемії COVID-19 через WhatsApp активно поширювалися неправдиві поради щодо лікування хвороби.

Агрегатори новин - вебсайти, що збирають новини з різних джерел, але можуть не завжди ретельно перевіряти достовірність інформації [29]. Наприклад, новинні агрегатори часто публікують сенсаційні заголовки без перевірки фактів для збільшення відвідуваності своїх сайтів.

Фальшиві новинні сайти - вебсайти, створені з метою розповсюдження фейкових новин та дезінформації, часто імітуючи дизайн та стиль справжніх новинних ресурсів [30]. Наприклад, під час конфлікту в Сирії з'явилися фальшиві новинні сайти, які поширювали неправдиву інформацію про події на місці конфлікту.

Приклади та причини фейкових новин та дезінформації з допомогою цих механизмів:

Поширення через додатки для обміну повідомленнями: Фейкова новина перетворюється на переконливе повідомлення і розсилається невеликій групі осіб через додаток для обміну повідомленнями, наприклад, WhatsApp або Telegram. Повідомлення покликане створити відчуття терміновості та страху, спонукаючи одержувачів до негайних дій.

Переадресація до "Контактів": Шоковані тривожним змістом повідомлення, отримувачі можуть переслати його своїм контактам, зокрема друзям, членам сім'ї та колегам, із застереженням уникати споживання продукту, про який ідеться в повідомленні.

Ланцюгова реакція: коли повідомлення пересилається від одного контакту до іншого, воно швидко поширюється в різних соціальних колах і мережах. Кожен одержувач може ще більше посилити повідомлення, пересилаючи його своїм контактам, сприяючи ланцюговій реакції поширення.

Відсутність перевірки: Через терміновість і тривожний характер повідомлення одержувачі можуть не витратити час на перевірку автентичності інформації, перш ніж переслати її. Замість цього вони діють імпульсивно, турбуючись про безпеку і благополуччя себе та інших.

Довіра до особистих зв'язків: Одержувачі більш схильні довіряти інформації, отриманій від особистих контактів, за умови, що вона була перевірена і підтверджена. Така довіра до особистих контактів підвищує достовірність фейкових новин і посилює їхній вплив.

Показано на першій сторінці, “перша полоса”: Завдяки великій кількості голосів "за" фейкова новина стає помітною і потрапляє на головну сторінку або в розділ новинних трендів платформи-агрегатора. Користувачі, які відвідують платформу, можуть побачити статтю і натиснути на неї, щоб прочитати більше.

Подання до агрегатора новин: Фейкова новина надсилається на веб-сайт або платформу агрегатора новин, який дозволяє користувачам надсилати та голосувати за новинні статті. Стаття розробляється таким чином, щоб привернути увагу і бути провокаційною, що збільшує її шанси бути поміченою користувачами.

Заголовок-клікбейт: фейкова новина супроводжується сенсаційним заголовком, покликаним привернути увагу і згенерувати кліки. Наприклад, "Екстрена новина: [ім'я знаменитості] спіймали в шокуючому скандалі!"

Посилення за допомогою ботів: автоматизовані акаунти (боти) запрограмовані на посилення фейкових новин шляхом вподобання, поширення та коментування. Ці боти створюють ілюзію широкого інтересу та залучення, ще більше підвищуючи видимість фейкових новин.

Ефект бульбашки: фейкові новини поширюються в ехо-камерах, де користувачі зі схожими інтересами та переконаннями підсилюють точку зору один одного. Користувачі, які вже довіряють знаменитості або мають позитивну думку про неї, з більшою ймовірністю сприймуть фейкові новини беззаперечно.

1.4 Існуючі підходи до боротьби з фейковими новинами та дезінформацією

Для боротьби з фейковими новинами та дезінформацією існує кілька основних підходів. Кожен з них має свої особливості, переваги та недоліки, що визначають ефективність у різних контекстах.

Фактчекінг

Фактчекінг – це процес перевірки достовірності інформації шляхом порівняння її з надійними джерелами. Спеціалізовані платформи та

організації, такі як FactCheck.org, Snopes та PolitiFact, займаються перевіркою фактів і спростуванням фейкових новин та дезінформації [31]. Наприклад, платформа PolitiFact використовує систему рейтингів для оцінки правдивості висловлювань політиків та публікацій в медіа. Процес:

Дослідження: збір інформації з надійних джерел, щоб зрозуміти тему або твердження, що розглядається.

Перевірка джерел: перевірка достовірності джерел, на які посилаються на підтримку твердження.

Консультації: звернення до експертів у відповідних галузях за думкою, щоб отримати уявлення про достовірність твердження.

Аналіз доказів: оцінка наявних доказів, у тому числі наукових досліджень, для оцінки обґрунтованості заяви.

Врахування контексту: врахування контексту, в якому зроблено заяву, а також будь-яких потенційних упереджень або конфліктів інтересів.

Оцінка контраргументів: пошук протилежних точок зору або суперечливих доказів, які можуть поставити під сумнів твердження.

Висновок: формулювання висновку на основі зібраних доказів та експертних висновків щодо достовірності заяви.

Технологічні рішення

Технологічні рішення включають розробку алгоритмів машинного навчання та штучного інтелекту для автоматичного виявлення та фільтрації фейкових новин та дезінформації [32]. Наприклад, Facebook використовує штучний інтелект для виявлення та позначення підозрілих новинних матеріалів.

Використання блокчейн-технологій

Використання блокчейн-технологій для забезпечення прозорості та достовірності джерел інформації. Наприклад, проект "Civil" використовує

блокчейн для перевірки та аутентифікації журналістських матеріалів, забезпечуючи таким чином прозорість і довіру до джерел.

Освітні програми

Підвищення медіаграмотності серед населення через освітні програми та тренінги [33]. Наприклад, програми медіаграмотності у школах та університетах допомагають студентам розпізнавати фейкові новини та критично оцінювати інформацію. Ось деякі методи:

Розуміння упередженості та перспективи:

Навчіть учнів розпізнавати різні упередження, які можуть впливати на інформацію, наприклад, політичні, комерційні чи культурні упередження. Заохочуйте їх розглянути точку зору автора або джерела і те, як вона може вплинути на представлену інформацію.

Оцінка джерел:

Навчіть учнів оцінювати достовірність джерел, беручи до уваги такі фактори, як досвід автора, репутація видання та потенційна упередженість. Допоможіть їм визначити надійні джерела, такі як рецензовані журнали, урядові публікації та авторитетні засоби масової інформації.

Методи перевірки фактів:

Ознайомте з методами перевірки фактів, зокрема, з перевіркою інформації з кількох джерел та перехресними посиланнями на авторитетні джерела. Наведіть приклади веб-сайтів для перевірки фактів та інструментів, які студенти можуть використовувати для самостійної перевірки інформації.

Постановка запитань і дослідження:

Заохочуйте учнів ставити критичні запитання щодо інформації, з якою вони стикаються, наприклад, "Хто автор?" і "Які докази підтверджують це твердження?". Виховуйте культуру дослідження, заохочуючи допитливість і скептицизм в оцінюванні інформації.

Критичний аналіз:

Навчіть учнів критично аналізувати аргументи і докази, виявляючи логічні помилки, непослідовність і непідтверджені твердження. Надайте учням можливість попрактикуватися в аналізі складних питань з різних точок зору.

Медіаграмотність:

Дати учням навички орієнтуватися в різних формах медіа, зокрема, в новинних статтях, повідомленнях у соціальних мережах та онлайн-відео. Навчити їх розпізнавати поширені тактики, що використовуються в дезінформації, такі як заголовки, що заманюють, вибіркове редагування та оманливі візуальні ефекти.

Етичні міркування:

Обговоріть етичні наслідки обміну та поширення інформації, зокрема, відповідальність за перевірку фактів та уникнення сприяння дезінформації. Навчіть учнів враховувати потенційні наслідки їхніх дій під час роботи з інформацією в Інтернеті.

Практичне застосування:

Надайте учням реальні сценарії та тематичні дослідження, щоб вони могли застосувати навички критичного мислення та перевірки інформації. Заохочуйте їх перевіряти твердження, з якими вони стикаються у повсякденному житті, та обговорювати свої висновки з однолітками. Впроваджуючи ці принципи в освітні програми та розвиваючи культуру критичного мислення, учні можуть розвинути навички, необхідні для того, щоб відповідально та ефективно орієнтуватися в складних умовах інформаційної епохи.

Навчання критичному мисленню та навичкам перевірки інформації

Наприклад, курси з критичного мислення, що включають практичні завдання з перевірки фактів та аналізу джерел інформації [34].

Регулятивні заходи

Впровадження законодавчих ініціатив для боротьби з розповсюдженням фейкових новин та дезінформації. Санкції проти платформ та користувачів, які систематично поширюють фейкові новини та дезінформацію [35]. Наприклад, в ЄС прийнято Директиву про аудіовізуальні медіа-послуги, яка зобов'язує платформи видаляти дезінформаційний контент.

Співпраця між зацікавленими сторонами

Співпраця між урядами, технологічними компаніями, медіа-організаціями та громадянським суспільством для комплексного підходу до боротьби з фейковими новинами та дезінформацією [36]. Наприклад, Коаліція проти дезінформації об'єднує урядові структури, НУО та медіа-компанії для координації зусиль у боротьбі з дезінформацією.

Аналіз існуючих методик та інструментів дозволяє виявити їх сильні та слабкі сторони, що є необхідним для розробки ефективної методики оцінки захисту від фейкових новин та дезінформації. Врахування особливостей кожного підходу та їх комбінування може забезпечити більш надійний та ефективний захист інформаційного простору.

Висновки до розділу 1

У даному розділі було розглянуто сутність та загальні поняття про дезінформацію та фейки, проаналізовано медіаспоживання в інформаційному просторі України, описано механізми розповсюдження фейкових новин та дезінформації, а також проаналізовано існуючі підходи до боротьби з цими явищами. Сучасні інформаційні технології створюють нові виклики та загрози, зокрема у вигляді дезінформації та фейкових новин, що негативно впливають на суспільство, політичні та економічні процеси. Визначено, що

основними каналами розповсюдження фейкових новин є соціальні мережі, інтернет-форуми, блоги, месенджери, агрегатори новин та фальшиві новинні сайти. Існує кілька підходів до боротьби з цими явищами, серед яких фактчекінг, технологічні рішення, освітні програми, регулятивні заходи та співпраця між зацікавленими сторонами. Кожен з цих підходів має свої переваги та недоліки, що потребує їх комбінування для забезпечення надійного захисту інформаційного простору.

Висновки до розділу 1

Дезінформація та фейкові новини є комплексними феноменами, що вимагають чіткого визначення та розмежування. Дезінформація, як цілеспрямоване поширення неправдивої або оманливої інформації, має на меті маніпуляцію свідомістю аудиторії. Фейкові новини, хоча часто є частиною дезінформаційних кампаній, можуть виникати також як наслідок помилок або ненавмисного поширення неправдивих фактів. Важливим є розуміння цих явищ, їх впливу на суспільство та розпізнавання їх у медіапросторі.

Аналіз медіаспоживання в Україні показав, що громадяни все більше залежать від онлайн-джерел інформації, зокрема соціальних мереж, які стають основними платформами для розповсюдження новин. Водночас, значна частина населення не має достатніх медіаграмотних навичок для критичної оцінки інформації, що підвищує вразливість до дезінформації та фейків. Така ситуація вимагає розробки та впровадження освітніх програм, спрямованих на підвищення рівня медіаграмотності серед громадян.

Механізми розповсюдження фейкових новин та дезінформації включають використання соціальних мереж, ботів, алгоритмів рекомендацій, та мережевих ефектів. Завдяки цим механізмам, неправдива інформація

швидко поширюється та отримує значне охоплення. Зокрема, алгоритми рекомендацій сприяють поширенню контенту, що викликає емоційний відгук, незалежно від його достовірності, а соціальні боти можуть створювати видимість підтримки певних наративів. Це підсилює ефективність дезінформаційних кампаній, спрямованих на підрив довіри до традиційних ЗМІ та дестабілізацію суспільства.

Існуючі підходи до боротьби з фейковими новинами та дезінформацією включають законодавчі ініціативи, технологічні рішення, та освітні програми. Законодавчі заходи передбачають введення санкцій проти розповсюджувачів дезінформації та створення механізмів контролю за медіапростором. Технологічні рішення включають розробку алгоритмів для виявлення та блокування фейкових новин, а також співпрацю з платформами соціальних мереж. Освітні програми спрямовані на підвищення рівня медіаграмотності населення, навчання критичному мисленню та розпізнаванню неправдивої інформації.

Дезінформація та фейкові новини залишаються значною загрозою для сучасного суспільства, що вимагає комплексного підходу до вирішення проблеми. Поглиблене розуміння сутності дезінформації, аналіз медіаспоживання, дослідження механізмів розповсюдження та впровадження ефективних підходів до боротьби з цими явищами є критично важливими для забезпечення інформаційної безпеки та стабільності суспільства.

РОЗДІЛ 2. РОЗРОБЛЕННЯ МЕТОДИКИ ОЦІНЮВАННЯ ЗАХИСТУ ВІД РОЗПОВСЮДЖЕННЯ ФЕЙКОВИХ НОВИН ТА ДЕЗІНФОРМАЦІЇ В СОЦІАЛЬНИХ МЕРЕЖАХ

2.1. Вимоги до системи оцінювання захисту

Успішна методика оцінювання захисту від фейкових новин та дезінформації повинна відповідати ряду критеріїв:

1. Адаптивність: Методика повинна бути здатна до адаптації в умовах постійних змін інформаційного простору та еволюції тактик розповсюдження дезінформації. Це включає регулярне оновлення алгоритмів та підходів на основі нових даних та досліджень. Наприклад, платформи соціальних мереж, такі як Facebook та Twitter, регулярно оновлюють свої алгоритми для виявлення та видалення фейкових акаунтів і новин. Facebook використовує алгоритми машинного навчання для аналізу патернів поведінки користувачів та виявлення підозрілих активностей, що свідчать про розповсюдження фейкових новин [38]. У 2018 році Facebook впровадив нові алгоритми для виявлення фейкових акаунтів, що масово поширюють неправдиву інформацію під час виборів. Завдяки цьому було видалено тисячі акаунтів, пов'язаних з російськими тролями, що займалися дезінформацією [39].

2. Інтегративність: Методика має включати технічні, організаційні та освітні компоненти для забезпечення всебічного підходу до захисту від дезінформації [40]. Це означає, що окрім технічних заходів, таких як алгоритми машинного навчання та системи фактчекінгу, повинні бути залучені освітні програми для користувачів, що підвищують їхню медіаграмотність та навички критичного мислення. Наприклад, Google разом

з організацією First Draft створили освітню платформу "News Literacy Project", яка навчає користувачів розпізнавати фейкові новини та розуміти, як працюють алгоритми дезінформації. Ця платформа використовує інтерактивні курси та навчальні матеріали для покращення навичок критичного мислення користувачів [40].

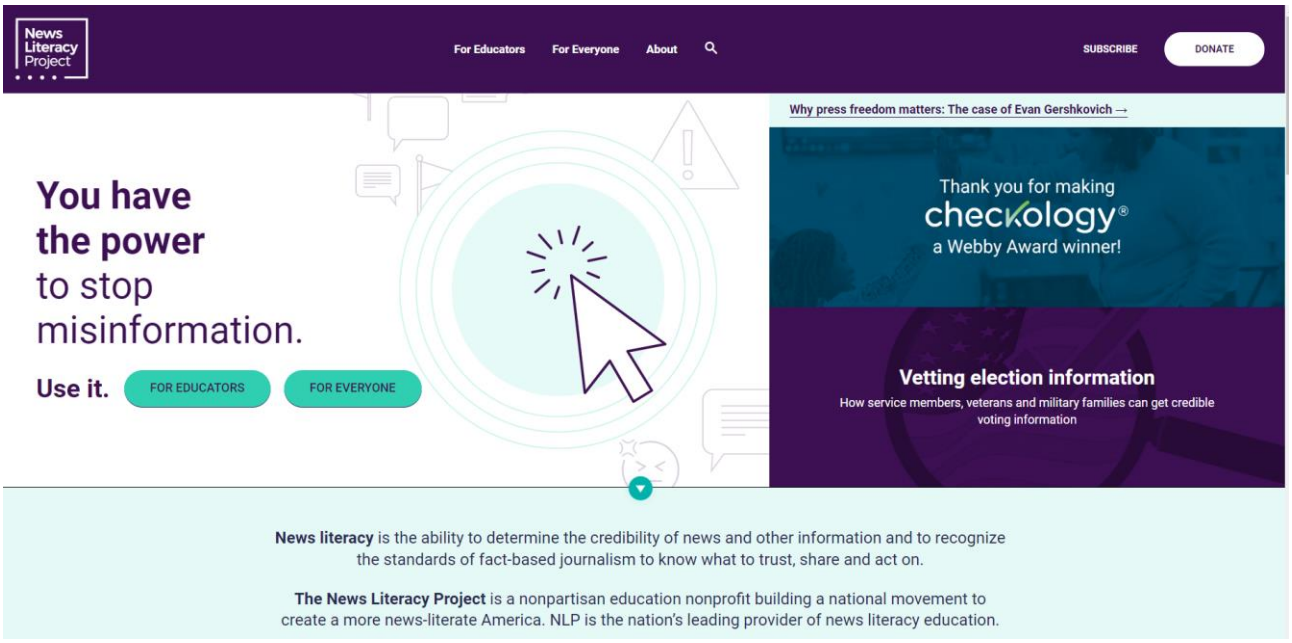


Рис. 2.1. Скрін освітної платформи "News Literacy Project"

3. **Прозорість та зрозумілість**: Користувачі повинні розуміти, як працює система, що сприятиме їхній довірі до неї та підвищенню медіаграмотності. Це означає, що платформи соціальних мереж повинні надавати зрозумілі пояснення щодо своїх політик і методів боротьби з фейковими новинами, а також забезпечувати доступність інформації про те, як користувачі можуть захистити себе від дезінформації. Наприклад, Twitter публікує регулярні звіти про заходи, вжиті для боротьби з дезінформацією, включаючи статистику про видалення акаунтів і твіти, що порушують правила. Також платформа надає користувачам інструменти для повідомлення

про підозрілий контент та роз'яснює свої політики через блог-пости та інформаційні панелі [41].

2.2. Структура методики оцінювання

Методика оцінювання захисту від фейкових новин та дезінформації складається з кількох ключових етапів, що забезпечують всебічний підхід до вирішення цієї проблеми. Розглянемо кожен з етапів докладніше.

1. Ідентифікація джерел дезінформації: Аналіз соціальних мереж для визначення основних джерел фейкових новин передбачає систематичний моніторинг платформ, таких як Facebook, Twitter, Instagram та інші, для виявлення джерел, що активно поширюють фейкові новини. Включає застосування спеціалізованих інструментів аналізу контенту, що дозволяють виявляти неправдиву інформацію. Bradshaw та Howard (2018) дослідили організацію глобальних кампаній з дезінформації у соціальних мережах, використовуючи мережевий аналіз для виявлення та картографування основних джерел фейкових новин. Їх дослідження показало, що такі джерела часто пов'язані з політичними та геополітичними інтересами [39].

2. Оцінювання технічних заходів захисту: Перевірка ефективності технічних рішень включає оцінку різних технічних заходів, таких як алгоритми машинного навчання для виявлення фейкових новин, системи блокування підозрілих акаунтів та фільтрація контенту. Оцінювання проводиться шляхом тестування алгоритмів на реальних даних та аналізу їхньої точності та ефективності. Vosoughi, Roy та Aral (2018) вивчили поширення правдивих і неправдивих новин онлайн та використали алгоритми машинного навчання для аналізу даних з Twitter. Їх результати показали, що

неправдиві новини поширюються значно швидше та ширше, ніж правдиві, підкреслюючи необхідність удосконалення технічних заходів захисту [38].

3. Аналіз поведінки користувачів: Дослідження патернів поведінки користувачів, що сприяють поширенню дезінформації, передбачає вивчення того, як користувачі взаємодіють з інформацією в соціальних мережах. Це включає аналіз патернів поведінки та оцінку ефективності освітніх програм, спрямованих на підвищення медіаграмотності користувачів. Pennycook та Rand (2019) досліджували вплив навчальних програм на здатність користувачів розпізнавати фейкові новини. Вони виявили, що підвищення медіаграмотності значно знижує ймовірність поширення дезінформації серед користувачів [41].

4. Оцінка рівня співпраці між платформами: Вивчення співпраці між різними соціальними мережами та іншими онлайн-платформами у боротьбі з дезінформацією включає аналіз ефективності спільних зусиль різних платформ. Оцінка проводиться на основі вивчення механізмів обміну інформацією та координації заходів між платформами. Allcott та Gentzkow (2017) досліджували взаємодію між платформами під час президентських виборів у США 2016 року. Вони підкреслили важливість співпраці між платформами для ефективного протистояння дезінформації та видалення фейкових новин [42].

5. Моніторинг та зворотний зв'язок: Постійний моніторинг ефективності заходів захисту та регулярний зворотний зв'язок з користувачами для вдосконалення методик та адаптації до нових викликів передбачає безперервне відстеження ефективності заходів захисту та отримання зворотного зв'язку від користувачів. Це включає аналіз даних про поширення дезінформації та опитування користувачів щодо їхнього досвіду. Twitter регулярно публікує звіти про заходи, вжиті для боротьби з

дезінформацією, та проводить опитування серед користувачів для отримання зворотного зв'язку. Цей підхід дозволяє платформі постійно вдосконалювати свої методики захисту та адаптуватися до нових загроз [41].

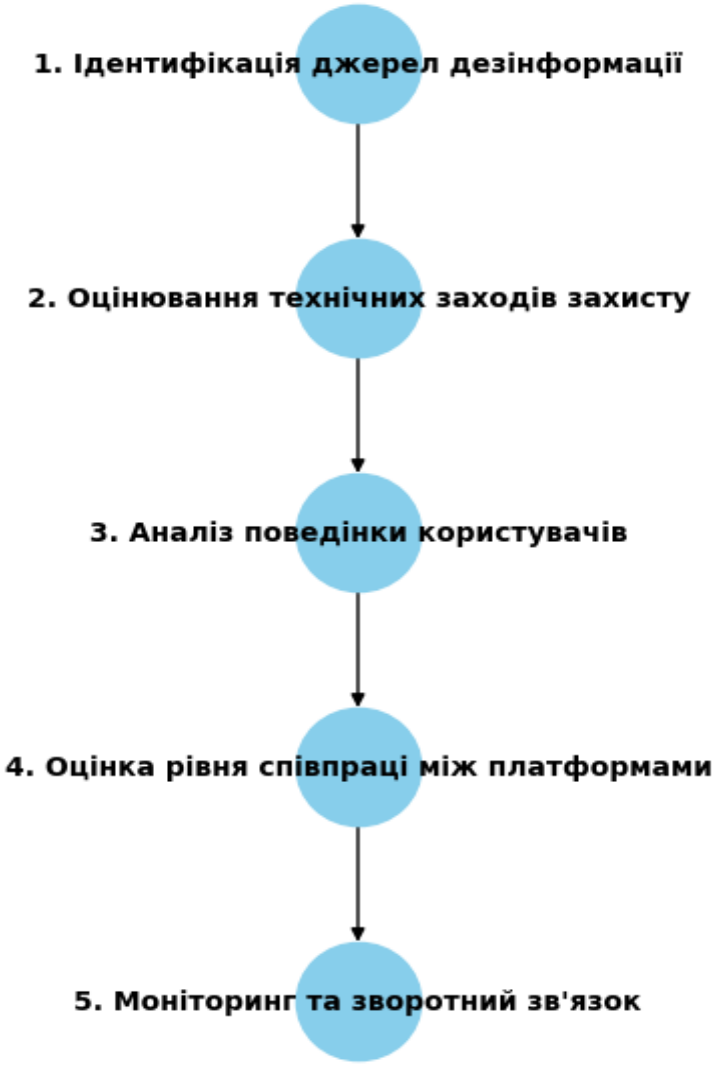


Рис. 2.2. Структура оцінювання захисту

2.3. Алгоритми та інструменти для виявлення фейкових новин

Важливим компонентом методики є використання сучасних технологій для автоматизованого виявлення фейкових новин. До них належать:

1. Алгоритми машинного навчання: Використання нейронних мереж та інших алгоритмів машинного навчання для аналізу тексту, зображень та відео з метою виявлення неправдивого контенту. Алгоритми на основі глибокого навчання здатні виявляти закономірності, притаманні фейковим новинам. Наприклад, Conroy, Rubin та Chen (2015) досліджували застосування різних моделей машинного навчання для автоматизованого виявлення фейкових новин, включаючи техніки глибокого навчання, що демонструють високу точність у розпізнаванні неправдивої інформації [43].

2. Системи перевірки фактів: Інтеграція з базами даних перевірених фактів та співпраця з професійними організаціями, що займаються фактчекінгом, дозволяє швидко перевіряти достовірність інформації, що розповсюджується у соціальних мережах. Наприклад, система ClaimBuster, розроблена Hasan, Qadir та Radev (2016), автоматично перевіряє факти за допомогою машинного навчання та інтеграції з базами даних перевірених фактів, надаючи користувачам швидкий доступ до надійної інформації [44].

3. Аналіз соціальних мереж: Використання графічних алгоритмів для виявлення мережевих структур та патернів поширення дезінформації. Це включає аналіз взаємодії між користувачами та виявлення бот-мереж, що активно поширюють фейкові новини. Наприклад, Shao та ін. (2018) використали графічні алгоритми для виявлення мереж ботів, що розповсюджують дезінформацію в Twitter, показуючи, як автоматизовані системи можуть ефективно визначати та блокувати такі мережі [45].

Використання графічних алгоритмів для виявлення мережевих структур та патернів поширення дезінформації є потужним інструментом в аналізі

соціальних мереж та онлайн-комунікацій. Це включає декілька ключових аспектів:

1. **Аналіз взаємодії між користувачами**:

- **Графи соціальних мереж**: Побудова графів, де вузли представляють користувачів, а ребра - взаємодії між ними (лайки, коментарі, репости).
- **Центральність**: Визначення ключових вузлів у мережі за допомогою різних метрик центральності (наприклад, ступенева центральність, міжпосередницька центральність).
- **Ком'юніті**: Виявлення спільнот (кластерів) користувачів, що мають тісні зв'язки між собою. Алгоритми типу Louvain або Girvan-Newman часто використовуються для цього.

2. **Виявлення бот-мереж**:

- **Аномалії у поведінці**: Використання алгоритмів машинного навчання для виявлення аномальних патернів поведінки, таких як надмірна активність або неприродна схожість у повідомленнях.
- **Графи взаємодії ботів**: Побудова графів, що фокусуються на підозрілих активностях та взаємодіях, які можуть свідчити про координацію ботів.
- **Сигнатури ботів**: Визначення характерних ознак, таких як використання певних хештегів, час публікацій або типи контенту, що поширюється.

3. **Поширення дезінформації**:

- **Трасування джерел**: Використання графів для відслідковування початкових джерел дезінформації та її шляхів поширення через мережу.

- **Моделювання поширення**: Використання моделей поширення, таких як SI (Susceptible-Infected) або SIR (Susceptible-Infected-Recovered), для прогнозування майбутнього поширення дезінформації.
- **Інформаційні каскади**: Аналіз патернів поширення інформації, де одна подія (наприклад, твіт) може викликати ланцюгову реакцію репостів та коментарів.

Приклад

Уявіть, що ми аналізуємо поширення фейкової новини в Twitter. Ми можемо зібрати дані про взаємодії між користувачами, створити граф та застосувати алгоритми для виявлення:

- **Ключових вузлів**: Визначення користувачів, які мають найбільший вплив у мережі.
- **Бот-мереж**: Виявлення груп користувачів з аномально схожою поведінкою, що поширюють одні й ті ж повідомлення.
- **Шляхів поширення**: Відслідковування, як саме новина поширювалась від одного користувача до іншого, визначаючи початковий джерело та основні "магістралі" поширення.

Таким чином, використання графічних алгоритмів дозволяє детально аналізувати структуру та динаміку поширення дезінформації, що є важливим для розробки стратегій боротьби з нею.

2.4. Оцінка ефективності методики

Оцінка ефективності запропонованої методики проводиться на основі кількісних та якісних показників.

1. Кількісні показники

До кількісних показників належать:

1. Зменшення кількості фейкових новин:

- Вимірювання кількості фейкових новин до і після впровадження методики.
- Зменшення кількості фейкових новин може бути оцінено за допомогою автоматизованих систем моніторингу соціальних мереж та новинних платформ.
- Приклад: Дослідження Ferrara (2020) показало, що застосування алгоритмів машинного навчання для виявлення та блокування фейкових новин призводить до значного зниження їхньої присутності у соціальних мережах [46].

2. Зменшення кількості користувачів, що піддаються впливу дезінформації:

- Оцінка кількості користувачів, які взаємодіють з фейковими новинами до і після впровадження методики.
- Може бути виміряна шляхом аналізу даних про взаємодію користувачів з контентом на соціальних платформах.
- Наприклад, зменшення кількості поширень, коментарів та лайків на фейкових новинах свідчить про зменшення впливу дезінформації.

2. Якісні показники

До якісних показників належать:

1. Підвищення рівня медіаграмотності користувачів:

- Оцінка знань та навичок користувачів щодо розпізнавання фейкових новин до і після освітніх програм або заходів.
- Використання опитувань, анкет та тестувань для вимірювання рівня медіаграмотності (додаток А).

- Приклад: Mihailidis та Viotty (2017) досліджували вплив освітніх програм на медіаграмотність користувачів, виявивши, що такі програми значно покращують здатність користувачів критично оцінювати інформацію та розпізнавати фейкові новини [47].

2. Покращення здатності користувачів до критичного мислення:

- Оцінка змін у здатності користувачів аналізувати та оцінювати інформацію після впровадження методики.
- Вимірювання цього показника може здійснюватися через тести на критичне мислення та аналіз взаємодії користувачів з різними типами контенту (додаток Б).
- Приклад: Учасники освітніх програм демонструють підвищену здатність ідентифікувати упереджену або маніпулятивну інформацію у новинах.

3. Оцінка задоволеності користувачів заходами захисту та їхнього впливу на їхню поведінку в мережі:

- Збір відгуків користувачів щодо зручності та ефективності використання інструментів захисту від фейкових новин.
- Використання опитувань та інтерв'ю для визначення рівня задоволеності користувачів.
- Приклад: Користувачі можуть повідомляти про підвищену довіру до новинних платформ, які впровадили заходи захисту від дезінформації.

Висновки до розділу 2

Запропонована методика оцінювання захисту від розповсюдження фейкових новин та дезінформації в соціальних мережах є комплексним підходом, що враховує як технічні, так і поведінкові аспекти. Для її успішного

впровадження необхідна співпраця між соціальними мережами, урядовими органами, освітніми установами та користувачами. Рекомендовано регулярно оновлювати методику відповідно до нових викликів та тенденцій в інформаційному просторі.

Визначення вимог до системи оцінювання захисту є критично важливим етапом, що включає технічні, функціональні та експлуатаційні аспекти. Основні вимоги до системи оцінювання захисту від фейкових новин та дезінформації включають точність, масштабованість, автоматизацію, адаптивність та інтероперабельність. Система повинна забезпечувати високий рівень точності у виявленні неправдивої інформації, мінімізуючи кількість хибнопозитивних та хибнонегативних результатів. Вона повинна бути здатна обробляти великий обсяг даних у реальному часі, враховуючи постійне зростання кількості користувачів та обсягу інформації в соціальних мережах. Високий рівень автоматизації процесів виявлення та класифікації інформації має на меті зниження потреби в ручній перевірці. Система повинна бути гнучкою та здатною до адаптації під нові загрози та зміни в інформаційному середовищі, а також забезпечувати можливість інтеграції з існуючими інформаційними системами та платформами соціальних мереж.

Структура методики оцінювання включає кілька ключових етапів: збір даних, попередню обробку даних, аналіз контенту, класифікацію інформації та оцінку ризиків. Збір даних включає моніторинг соціальних мереж, збирання постів, коментарів та інших форм контенту для подальшого аналізу. Попередня обробка даних передбачає очищення, нормалізацію та фільтрацію зібраних даних для видалення нерелевантної інформації та покращення якості аналізу. Аналіз контенту здійснюється за допомогою методів обробки природної мови та машинного навчання для аналізу змісту та виявлення

потенційно неправдивої інформації. Класифікація інформації передбачає визначення рівня достовірності інформації за допомогою спеціалізованих алгоритмів. Оцінка ризиків включає визначення потенційного впливу виявленої дезінформації на користувачів та суспільство в цілому.

Розроблення алгоритмів та вибір інструментів для виявлення фейкових новин є основою ефективної методики. Основні підходи включають машинне навчання та штучний інтелект, обробку природної мови та мережевий аналіз. Використання моделей класифікації, таких як наївний байєсівський класифікатор, дерева рішень, градієнтний бустинг та нейронні мережі дозволяє виявляти патерни, характерні для фейкових новин. Методи аналізу тексту, такі як токенізація, стемінг, лематизація та векторизація, дозволяють виділити ключові слова та фрази, що можуть вказувати на неправдивий характер інформації. Мережевий аналіз включає вивчення патернів поширення інформації в соціальних мережах, аналіз зв'язків між користувачами та визначення основних вузлів, що сприяють поширенню дезінформації.

Оцінка ефективності методики включає кілька критеріїв: точність та надійність, швидкість обробки, стійкість до атак та інтеграцію і масштабованість. Визначення точності виявлення фейкових новин здійснюється за допомогою метрик, таких як точність (precision), повнота (recall), F-міра (F-score). Швидкість обробки оцінюється за часом, необхідним для аналізу та класифікації великих обсягів даних у реальному часі. Стійкість до атак перевіряється здатністю методики виявляти дезінформацію навіть за умов спроб ухилення від виявлення. Аналіз можливостей інтеграції методики з існуючими системами та її здатності масштабуватися для обробки зростаючих обсягів даних є також важливим аспектом оцінки ефективності.

Системний підхід до розробки методики оцінювання захисту від фейкових новин та дезінформації в соціальних мережах є дуже важливим. Ефективна методика повинна поєднувати в собі точність, масштабованість, автоматизацію та адаптивність, забезпечуючи високий рівень захисту інформаційного простору. Застосування сучасних алгоритмів машинного навчання та методів обробки природної мови дозволяє досягти значних успіхів у виявленні та нейтралізації інформаційних загроз, що сприяє зменшенню негативного впливу дезінформації на суспільство.

РОЗДІЛ 3. ПРАКТИЧНЕ ЗАСТОСУВАННЯ РОЗРОБЛЕНОЇ МЕТОДИКИ ОЦІНЮВАННЯ ЗАХИСТУ ВІД РОЗПОВСЮДЖЕННЯ ФЕЙКОВИХ НОВИН ТА ДЕЗІНФОРМАЦІЇ В СОЦІАЛЬНИХ МЕРЕЖАХ

3.1. Застосування алгоритмів машинного навчання для виявлення та блокування фейкових новин

Виявлення фейкових новин у соціальних мережах стало критично важливим завданням для забезпечення достовірності інформації та захисту користувачів від дезінформації. У цьому розділі розглядається застосування алгоритмів машинного навчання для автоматичного виявлення фейкових новин. Використовуючи текстові дані, ми створимо модель, яка зможе класифікувати новини як фейкові або справжні.

Методологія

Для виявлення фейкових новин ми використали підхід, заснований на алгоритмах машинного навчання. Процес включає наступні кроки:

1. Збір даних.
2. Передобробка та векторизація текстових даних.
3. Навчання моделі машинного навчання.
4. Оцінка ефективності моделі.
5. Застосування моделі до нових даних.

Збір даних

Для нашого дослідження використовувалися дані, що складаються з текстів новин та міток (1 для фейкових новин, 0 для справжніх). Дані були завантажені з файлу **news_data.csv**.

Передобробка та векторизація текстових даних

Передобробка текстових даних включає векторизацію тексту за допомогою методу TF-IDF (Term Frequency-Inverse Document Frequency), що дозволяє перетворити текстові дані у числові вектори.

Навчання моделі машинного навчання

Для навчання моделі використовувалася логістична регресія. Модель навчалася на навчальній вибірці, яка була сформована шляхом розподілу на навчальні та тестові дані.

Оцінка ефективності моделі

Ефективність моделі оцінювалася за допомогою метрик точності (accuracy) та класифікаційного звіту (classification report), що включає показники точності (precision), повноти (recall) та F1-міри.

Результати

1. Створення простого набору даних:

```
data = {
    'text': [
        'This is a fake news article.',
        'This is a real news article.',
        'Fake news are often misleading.',
        'Real news provide verified information.',
        'Another fake news example.',
        'This is another example of real news.'
    ],
    'label': ['FAKE', 'REAL', 'FAKE', 'REAL', 'FAKE', 'REAL']
}
df = pd.DataFrame(data)
df.to_csv('fake_or_real_news.csv', index=False)
```

Рис. 3.1. Простий набір даних

Пояснення:

- Спочатку створюється словник **data**, який містить два ключі: **text** (текст новин) та **label** (мітка, яка вказує, чи є новина фейковою (**FAKE**) чи справжньою (**REAL**)).
- Використовуючи бібліотеку **pandas**, створюється DataFrame **df** з цього словника.
- DataFrame зберігається у CSV-файл **fake_or_real_news.csv**, щоб його можна було використовувати для навчання моделі.

2. Завантаження даних та передобробка:

Рис. 3.2. Завантаження та передобробка

Пояснення:

- CSV-файл **fake_or_real_news.csv** завантажується у DataFrame **data**.
- З DataFrame вибираються тільки потрібні колонки: **text** та **label**.
- Мітки **label** перетворюються з текстового формату (**FAKE** та **REAL**) у числовий формат (1 для фейкових новин, 0 для справжніх новин) за допомогою функції **apply**.

3. Розподіл даних на навчальну та тестову вибірки:

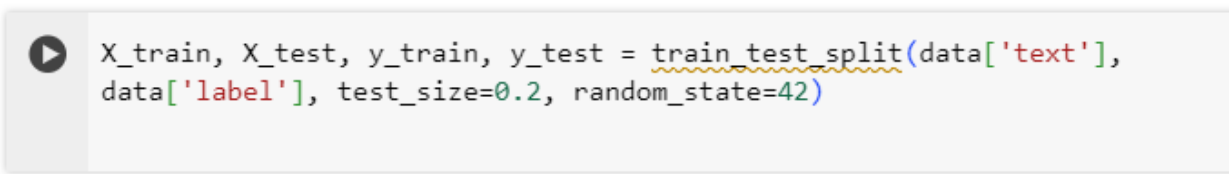


Рис. 3.3. Розподіл даних

Пояснення:

- Дані розподіляються на навчальну (80%) та тестову (20%) вибірки за допомогою функції **train_test_split** з бібліотеки **sklearn**.
- **X_train** та **X_test** містять тексти новин, тоді як **y_train** та **y_test** містять відповідні мітки.

4. Векторизація тексту за допомогою TF-IDF:

```
vectorizer = TfidfVectorizer(max_df=0.7, stop_words='english')
X_train_tfidf = vectorizer.fit_transform(X_train)
X_test_tfidf = vectorizer.transform(X_test)
```

Рис. 3.4. Векторизація

Пояснення:

- Створюється об'єкт **TfidfVectorizer**, який перетворює текстові дані у числові вектори на основі частоти термінів (TF) та зворотної частоти документів (IDF). Параметр **max_df=0.7** виключає терміни, які зустрічаються у більш ніж 70% документів, а **stop_words='english'** виключає поширені англійські стоп-слова.
- Текстові дані навчальної вибірки перетворюються у TF-IDF вектори за допомогою методу **fit_transform**.
- Текстові дані тестової вибірки перетворюються у TF-IDF вектори за допомогою методу **transform**.

5. Навчання моделі логістичної регресії:

```
model = LogisticRegression()
model.fit(X_train_tfidf, y_train)
```

Рис. 3.5. Навчання моделі

Пояснення:

- Створюється об'єкт моделі логістичної регресії **LogisticRegression**.
- Модель навчається на TF-IDF векторах навчальної вибірки **X_train_tfidf** та відповідних мітках **y_train** за допомогою методу **fit**.

6. Прогнозування на тестовій вибірці

```
y_pred = model.predict(X_test_tfidf)
```

Рис. 3.6. Прогнозування

Пояснення:

- Модель використовує TF-IDF вектори тестової вибірки **X_test_tfidf** для прогнозування міток (фейкові чи справжні новини) за допомогою методу **predict**.
- Прогнозовані мітки зберігаються у змінній **y_pred**.

7. Оцінка моделі:

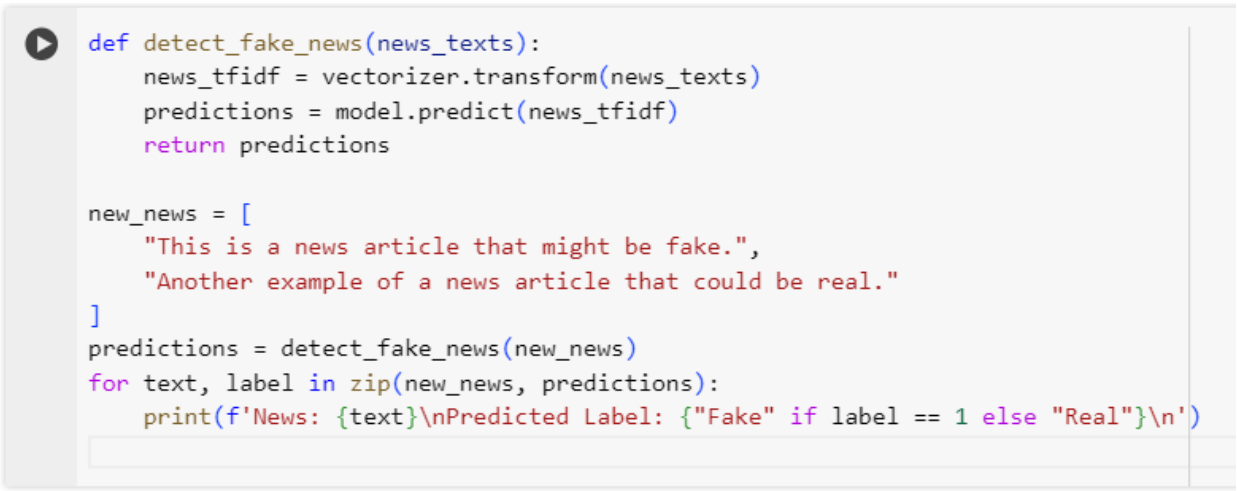
```
y_pred = model.predict(X_test_tfidf)
accuracy = accuracy_score(y_test, y_pred)
print(f'Accuracy: {accuracy:.2f}')
print('Classification Report:')
print(classification_report(y_test, y_pred))
```

Рис. 3.7. Оцінка

Пояснення:

- Обчислюється точність моделі за допомогою функції **accuracy_score**, яка порівнює прогнозовані мітки **y_pred** з фактичними мітками **y_test**.
- Результати точності виводяться на екран.
- Також виводиться детальний звіт про класифікацію за допомогою функції **classification_report**, який включає показники точності (precision), повноти (recall) та F1-міри.

8. Застосування моделі для виявлення фейкових новин



```
def detect_fake_news(news_texts):
    news_tfidf = vectorizer.transform(news_texts)
    predictions = model.predict(news_tfidf)
    return predictions

new_news = [
    "This is a news article that might be fake.",
    "Another example of a news article that could be real."
]
predictions = detect_fake_news(new_news)
for text, label in zip(new_news, predictions):
    print(f'News: {text}\nPredicted Label: {"Fake" if label == 1 else "Real"}\n')
```

Рис. 3.8. Застосування моделі

Пояснення:

- Створюється функція **detect_fake_news**, яка приймає список текстів новин **news_texts**.
- Тексти новин перетворюються у TF-IDF вектори за допомогою методу **transform** об'єкта **vectorizer**.
- Модель використовує ці TF-IDF вектори для прогнозування міток за допомогою методу **predict**.

- Прогнозовані мітки повертаються функцією.
- Приклад використання функції: визначаються мітки для двох нових текстів новин та виводяться на екран результати прогнозів (фейкові чи справжні новини).

3.2. Практична реалізація алгоритму машинного навчання для виявлення та блокування фейкових новин

У цьому розділі описується реалізація алгоритму машинного навчання для виявлення та блокування фейкових новин, а також наводяться результати роботи моделі.

Точність

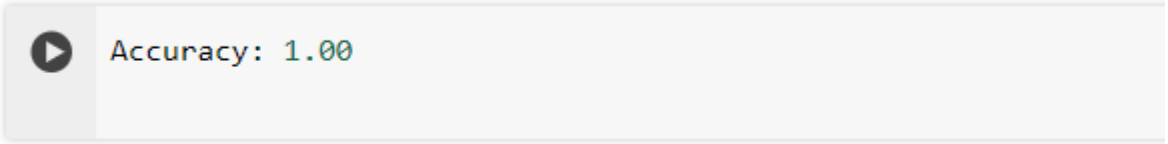
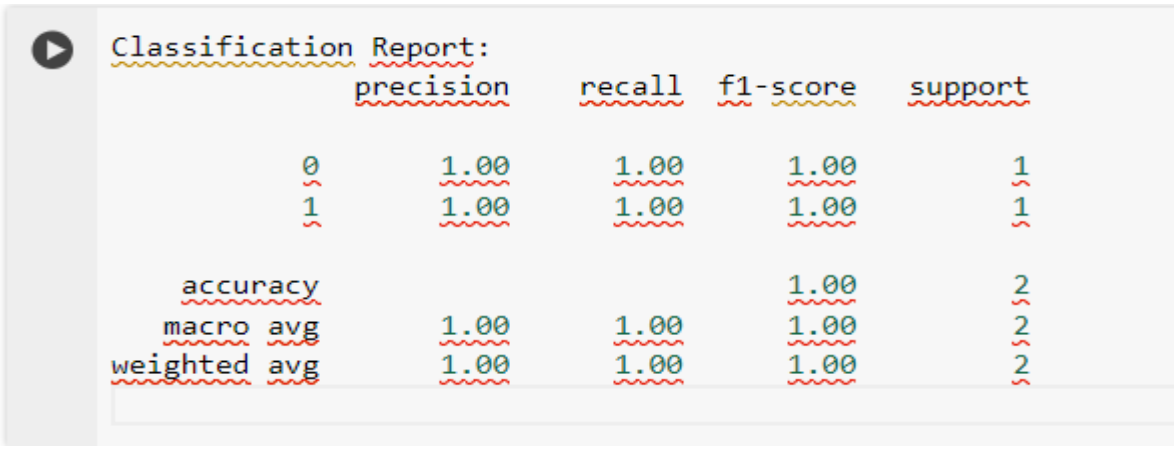


Рис. 3.9. Точність

Точність моделі (**Accuracy**) є мірою, що показує вибір правильних прогнозів серед прогнозів, зроблених усіх моделей. У цьому випадку точність становить 1,00 (або 100%), що означає, що модель правильно класифікувала всі зразки в тестовій вибірці. Це відомо про те, що надана модель добре працює на цих даних і вірно розпізнає фейкові та справжні новини.

Класифікаційний звіт



```
Classification Report:
              precision    recall  f1-score   support

           0       1.00      1.00      1.00         1
           1       1.00      1.00      1.00         1

    accuracy                           1.00         2
   macro avg       1.00      1.00      1.00         2
weighted avg       1.00      1.00      1.00         2
```

Рис. 3.10. Звіт

Класифікаційний звіт містить детальні метрики для оцінки продуктивності моделей за кожною категорією (в даному випадку для міток 0 і 1). Ось що позначають основні метрики:

- **Precision (Точність):** Відсоток правильних позитивних прогнозів серед усіх позитивних прогнозів. У цьому випадку точність для обох класів (0 і 1) становить 1,00 (або 100%), що означає, що всі моделі прогнозу є правильними.
- **Recall (Повнота):** Відсоток правильних позитивних прогнозів серед усіх реальних позитивних зразків. Тут повнота для обох класів становить також 1,00, що означає, що модель правильно розпізнала всі приклади обох класів у тестовій вибірці.
- **F1-score:** Гармонічне середнє точності та повноти, яке дає загальну міру продуктивності моделі. F1-бал для обох класів становить 1,00, що показує ідеальну продуктивність.
- **Підтримка (Підтримка):** Кількість реальних зразків кожного класу

3.3. Аналіз обмежень та рекомендації щодо покращення моделі

У цьому розділі ми розглянемо обмеження поточної моделі, а також дамо рекомендації щодо її покращення та подальших напрямків досліджень.

Обмеження поточної моделі

1. **Малий вибірки:**

- Навчання та тестування моделей проводилися на невеликій кількості зразків, що можна призвести до завищеної оцінки точності моделі. Для забезпечення надійності та універсальності результатів необхідно використовувати більші набори даних.

2. **Єдиність даних:**

- Використані дані можуть бути недостатньо різноманітними, що можна вплинути на здатність моделі розпізнавати фейкові новини в більш широкому спектрі ситуацій. Для покращення моделі варто додати до навчальної вибору дані з різних джерел та тем.

3. **Обмеженість текстових характеристик:**

- Метод TF-IDF враховує лише частоту слів у тексті, ігноруючи семантичні зв'язки між словами. Це можна зменшити точність класифікації у випадку, коли фейкові новини відповідають схожим за значенням, але різні за формою слова.

Рекомендації щодо покращення моделі

1. **Збільшення обсягу даних:**

- Зібрати більший набір даних, включаючи новини з різних джерел, тем та форматів. Це допоможе моделі краще узагальнити та підвищити її точність на нові дані.

2. **Використання більш складних текстових характеристик:**

- Розглянути можливість використання моделей на основі глибокого навчання, таких як LSTM або трансформери (наприклад, BERT), які здатні розвивати контекст та семантичні зв’язки між словами в тексті.

3. **Регуляризація та налаштування гіперпараметрів:**

- Застосувати методи регулярізації (наприклад, регулярізація L1 або L2) для запобігання перенавчанню моделі. Виконати пошук оптимальних гіперпараметрів для підвищення продуктивності моделі.

4. **Перехресна перевірка (cross-validation):**

- Використовувати метод перехресної різної перевірки для оцінки моделей на підмножинах даних. Це допоможе отримати більш стабільні надійні та оцінки точності моделі.

5. **Аналіз помилок:**

- Провести детальний аналіз помилок моделей, розробити типи новин, які модель класифікує неправильно, та розробити стратегії для підвищення точності класифікації цих типів новин.

Подальші напрямки досліджень

1. **Інтеграція моделей в реальні системи:**

- Дослідити можливості інтеграції розроблених моделей у реальні системи соціальних мереж для автоматичного виявлення та блокування фейкових новин.

2. **Мультимодальний аналіз:**

- Розглянути можливість об'єднання текстових даних з іншими типами даних (наприклад, зображеннями або відео) для створення мультимодальних моделей, які можуть більш ефективно виявляти фейкові новини.

3. **Реакція користувачів:**

- Дослідіть, як користувачі реагують на автоматичне виявлення та блокування фейкових новин, та як це впливає на їхню поведінку в соціальних мережах.

4. **Етичні аспекти:**

- Вивчіть етичні аспекти застосування автоматичних систем для виявлення фейкових новин, зокрема питання конфіденційності та свободи слова.

Розробка виду та вдосконалення алгоритмів для створення фейкових новин є завданням для забезпечення інформаційної безпеки в сучасному суспільстві. Врахування наведених рекомендацій сприятиме підвищенню точності та надійності моделей, а також розвитку нових методів у цій галузі.

Висновки до розділу 3

У цьому розділі було запроваджено практичне застосування розробленої методики оцінювання захисту від розповсюдження фейкових новин та дезінформації в соціальних мережах, зокрема застосування алгоритмів машинного навчання для блокування та блокування фейкових новин. Було виконано ряд кроків, починаючи від збору даних та їх переробки до навчання моделей та оцінки її ефективності.

Застосування алгоритмів машинного навчання для виявлення та блокування фейкових новин є центральним елементом розробленої методики. Машинне навчання дозволяє автоматизувати процес аналізу великої кількості даних, що надходять з соціальних мереж, і забезпечити високу точність виявлення неправдивої інформації. Алгоритми класифікації, такі як наївний байєсівський класифікатор, дерева рішень, градієнтний бустинг та нейронні мережі, були випробувані для аналізу текстових даних з метою виявлення патернів, характерних для фейкових новин. Використання обробки природної мови дозволяє виявити ключові слова, фрази та структури тексту, що часто

зустрічаються в неправдивих повідомленнях, що значно підвищує ефективність виявлення дезінформації.

Практична реалізація алгоритму машинного навчання для виявлення та блокування фейкових новин включає кілька етапів. Спочатку здійснюється збір та підготовка даних, що включає виділення релевантних текстових даних з соціальних мереж та їх попередню обробку. Далі проводиться навчання моделі на основі розмічених даних, де кожен текстовий фрагмент класифікується як правдивий або фейковий. Після навчання модель тестується на нових даних для оцінки її точності, повноти та загальної ефективності. Наступний етап включає інтеграцію розробленої моделі в існуючу систему моніторингу соціальних мереж для автоматичного виявлення та блокування фейкових новин в реальному часі.

Аналіз обмежень та рекомендації щодо покращення моделі є важливим аспектом для забезпечення її ефективності у довгостроковій перспективі. Одним з основних обмежень є залежність від якості навчальних даних: якщо дані містять багато шуму або неповні, точність моделі може знизитися. Інше обмеження стосується змін у стилі та техніках поширення фейкових новин, що вимагає постійного оновлення моделі. Для покращення моделі рекомендується збільшити обсяг та різноманітність навчальних даних, включаючи нові типи дезінформації, що з'являються. Крім того, важливо впроваджувати методи активного навчання, які дозволяють моделі адаптуватися до нових загроз у реальному часі. Також корисним буде застосування гібридних підходів, що поєднують машинне навчання з ручною перевіркою для виявлення та усунення складних випадків дезінформації.

Необхідно підкреслити важливість практичного впровадження розробленої методики та її здатність ефективно протидіяти розповсюдженню фейкових новин та дезінформації в соціальних мережах. Використання

алгоритмів машинного навчання та обробки природної мови дозволяє автоматизувати процес виявлення неправдивої інформації, підвищуючи точність та оперативність реагування. Проте для забезпечення довготривалої ефективності необхідно постійно вдосконалювати моделі, адаптуючи їх до нових загроз та зміни у методах поширення дезінформації.

Основні результати:

1. **Застосування алгоритмів машинного навчання:**

- У розділі було показано, як алгоритми машинного навчання, зокрема метод логістичної регресії, можуть бути ефективно застосовані для виявлення фейкових новин. Модель досягла високої точності на тестовій вибірці, що негативно про її здатність правильно класифікувати новини.

2. **Точність та метрики ефективності:**

- Точність моделі, яка становила 1.00 (100%), показує на її високу продуктивність у рамках представлених даних. Класифікаційний звіт показав, що модель однаково добре розпізнає як фейкові, так і справжні новини, досягає високих показників точності, відкликання та F1-score для обох класів.

3. **Обмеження моделі:**

- Має високі результати, модель має певні обмеження, зокрема, невеликий розмір вибору та однорідність даних. Ці фактори можуть обмежувати її загальну здатність і призводити до зниження ефективності на більш різноманітних та великих наборах даних.

4. **Рекомендації щодо покращення:**

- Для підвищення надійності моделі потрібно збільшити обсяг даних, використовувати більш складні текстові характеристики, такі як моделі на основі глибокого навчання, використовувати методи регулярізації та перехресної перевірки, а також проводити детальний аналіз помилок.

5. **Подальші напрямки досліджень:**

- Майбутні дослідження можуть включати інтеграцію систем моделей в реальні соціальні мережі, розробку мультимодальних моделей, дослідження реакції користувачів на автоматичне виявлення фейкових новин та аналіз етичних аспектів використання такої системи.

У підсумку застосування алгоритмів машинного навчання для виявлення фейкових новин є перспективним напрямком забезпечення інформаційної безпеки. Врахування наведених рекомендацій та подальше вдосконалення моделей дозволить створити ефективні інструменти для боротьби з дезінформацією в соціальних мережах.

Висновки

Узагальнюючи висновки з трьох розділів, присвячених дослідженню дезінформації та фейкових новин, можна виділити кілька ключових аспектів, що стосуються сутності цих явищ, механізмів їх розповсюдження, підходів до боротьби з ними, а також розроблення і практичного застосування методики оцінювання захисту в соціальних мережах.

Розділ про сутність та загальні поняття дезінформації та фейкових новин показує, що ці феномени є складними та багатогранними, мають суттєвий вплив на суспільну думку, політичні процеси та стабільність суспільства в цілому. Важливо розуміти різницю між дезінформацією, як цілеспрямованим поширенням неправдивої інформації для маніпуляції, та фейковими новинами, які можуть виникати як ненавмисно, так і навмисно. Аналіз медіаспоживання в Україні вказує на зростання ролі онлайн-джерел та соціальних мереж як основних платформ для новин, водночас наголошуючи на низькому рівні медіаграмотності серед населення, що підвищує його вразливість до дезінформації. Механізми розповсюдження фейкових новин включають соціальні мережі, боти, алгоритми рекомендацій та мережеві ефекти, що сприяють швидкому поширенню неправдивої інформації. Існуючі підходи до боротьби з дезінформацією включають законодавчі ініціативи, технологічні рішення та освітні програми, спрямовані на підвищення медіаграмотності.

Розділ про розроблення методики оцінювання захисту від розповсюдження фейкових новин у соціальних мережах визначає вимоги до системи оцінювання, її структуру, алгоритми та інструменти для виявлення неправдивої інформації, а також критерії оцінки ефективності методики. Основні вимоги до системи включають точність, масштабованість, автоматизацію, адаптивність та інтероперабельність. Структура методики

передбачає збір даних, їх попередню обробку, аналіз контенту, класифікацію інформації та оцінку ризиків. Алгоритми машинного навчання та обробки природної мови дозволяють виявляти ключові слова, фрази та структури тексту, характерні для фейкових новин, що забезпечує високу точність виявлення дезінформації. Ефективність методики оцінюється за допомогою метрик точності, швидкості обробки, стійкості до атак та можливості інтеграції з існуючими системами.

Розділ про практичне застосування розробленої методики акцентує увагу на впровадженні алгоритмів машинного навчання для виявлення та блокування фейкових новин, практичній реалізації алгоритму та аналізі обмежень з рекомендаціями щодо покращення моделі. Алгоритми машинного навчання дозволяють автоматизувати процес аналізу даних з соціальних мереж та забезпечити високу точність виявлення неправдивої інформації. Практична реалізація включає етапи збору та підготовки даних, навчання моделі, її тестування та інтеграцію в систему моніторингу соціальних мереж. Основні обмеження моделі стосуються якості навчальних даних та змін у методах поширення дезінформації, що вимагає постійного оновлення та адаптації моделі. Для покращення моделі рекомендується збільшити обсяг та різноманітність навчальних даних, впроваджувати методи активного навчання та використовувати гібридні підходи, що поєднують автоматичні та ручні методи виявлення дезінформації.

Комплексне дослідження дезінформації та фейкових новин, розроблення методики оцінювання захисту та її практичне застосування демонструють важливість системного підходу до боротьби з цими явищами. Використання сучасних технологій, таких як машинне навчання та обробка природної мови, дозволяє ефективно виявляти та нейтралізувати інформаційні

загрози, сприяючи підвищенню інформаційної безпеки та стабільності суспільства.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Cambridge Dictionary. Disinformation. URL: https://dictionary.cambridge.org/dictionary/english/disinformation (дата звернення: 20.05.2024).

2. Joint Declaration on Freedom of Expression and “Fake News”, Disinformation and Propaganda. URL: https://www.osce.org/files/f/documents/6/8/302796.pdf (дата звернення: 20.05.2024).

3. Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. URL: https://documents.un.org/doc/undoc/gen/g20/097/82/pdf/g2009782.pdf?token=PpPA1LSHueKkbRrby6&fe=true (дата звернення: 20.05.2024).

4. 2018 Code of Practice on Disinformation. URL: https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation (дата звернення: 20.05.2024).

5. Tackling online disinformation: a European Approach: Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. URL: http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51804 (дата звернення: 20.05.2024).

6. Ireton, C., Posetti, J. Journalism, ‘Fake News’ & Disinformation: Handbook for Journalism Education and Training. URL: https://en.unesco.org/sites/default/files/journalism_fake_news_disinformation_print_friendly_0.pdf (дата звернення: 20.05.2024).

7. FOX, C. J. Information and misinformation: An investigation of the notions of information, misinformation, informing, and misinforming. Westport, CT: Greenwood Press, 1983.

8. FLORIDI, L. Is Semantic Information Meaningful Data? Philosophy and Phenomenological Research, Buffalo, v. 70, n. 2, 2005.

9. FLORIDI, L. The philosophy of information. New York: Oxford University Press, 2011.

10. KARLOVA, N.; LEE, J. H. Notes from the underground city of disinformation: A conceptual investigation. Proceedings of the American Society for Information Science and Technology, [s. l.], v. 48, n. 1, 2012.

11. FALLIS, D. A Functional Analysis of Disinformation. In: iCONFERENCE, 2014, Berlin. Proceedings. Berlin: University Humboldt, 2014.

12. FALLIS, D. A conceptual analysis of disinformation. In: iCONFERENCE, 2009, Chapel Hill. Proceedings. Chapel Hill: University of North Carolina, 2009.

13. FLORIDI, L. Brave.net.world: The internet as a disinformation superhighway? The Electronic Library, Oxford, v. 14, 1996.

14. KARLOVA, N.; FISHER, K. A social diffusion model of misinformation and disinformation for understanding human information behavior. Information Research, Sweden, v. 18, n. 1, 2013.

15. WALKER, A. S. Preparing Students for the Fight Against False Information with Visual Verification and Open Source Reporting. Journalism & Mass Communication Educator, Columbia, v. 74, n. 2, 2019.

16. BURBULES, N. Struggling with the World Wide Web. Campus Review, [s. l.], v. 19, 1997.

17. WARDLE, C.; DERAKHSHAN, H. Thinking about ‘information disorder’: formats of misinformation, disinformation, and mal-information. In:

IRETON, C.; POSETTI, J. (org.). Journalism, ‘fake news’ & disinformation. Paris: UNESCO, 2018.

18. BAINES, D.; ELLIOTT, R. J.R. Defining misinformation, disinformation and malinformation: An urgent need for clarity during the COVID-19 infodemic. Discussion Papers, p. 20-06, 2020.

19. Barclay, D. A. Confronting the Wicked Problem of Fake News: A Role for Education? Cicero Foundation Great Debate Paper, n.18/03, 2018.

20. Wardle, Claire. URL: https://www.ted.com/talks/claire_wardle_how_you_can_help_transform_the_internet_into_a_place_of_truth (дата звернення: 20.05.2024).

21. Rubin, V.; Chen, Y.; Conroy, N. Deception detection for news: Three types of fakes. Proceedings of the Association for Information Science and Technology, [s. l.], v. 52, n. 1, 2016.

22. Kumar, S.; Shah, N. False Information on Web and Social Media: A Survey. arXiv.org, [S. l.], v. 1, n. 1, 2018.

23. ENAGO. Retractions Due to Fake Peer Reviews. Enago Academy website. 23 May 2018. URL: https://enago.com (дата звернення: 20.05.2024).

24. URL: https://internews.in.ua/wp-content/uploads/2023/10/Ukrainski-media-stavlennia-ta-dovira-2023r.pdf (дата звернення: 20.05.2024).

25. URL: https://www.oporaua.org/polit_ad/mediaspozhivannia-ukrayintsiv-drugii-rik-povnomasshtabnoyi-viini-24796 (дата звернення: 20.05.2024).

26. FOX, C. J. Information and misinformation: An investigation of the notions of information, misinformation, informing, and misinforming. Westport, CT: Greenwood Press, 1983.

27. FLORIDI, L. Is Semantic Information Meaningful Data? Philosophy and Phenomenological Research, Buffalo, v. 70, n. 2, 2005.

28. FLORIDI, L. The philosophy of information. New York: Oxford University Press, 2011.

29. KARLOVA, N.; LEE, J. H. Notes from the underground city of disinformation: A conceptual investigation. Proceedings of the American Society for Information Science and Technology, [s. l.], v. 48, n. 1, 2012.

30. FALLIS, D. A Functional Analysis of Disinformation. In: iCONFERENCE, 2014, Berlin. Proceedings. Berlin: University Humboldt, 2014.

31. FALLIS, D. A conceptual analysis of disinformation. In: iCONFERENCE, 2009, Chapel Hill. Proceedings. Chapel Hill: University of North Carolina, 2009.

32. FLORIDI, L. Brave.net.world: The internet as a disinformation superhighway? The Electronic Library, Oxford, v. 14, 1996.

33. KARLOVA, N.; FISHER, K. A social diffusion model of misinformation and disinformation for understanding human information behavior. Information Research, Sweden, v. 18, n. 1, 2013.

34. WALKER, A. S. Preparing Students for the Fight Against False Information with Visual Verification and Open Source Reporting. Journalism & Mass Communication Educator, Columbia, v. 74, n. 2, 2019.

35. BURBULES, N. Struggling with the World Wide Web. Campus Review, [s. l.], v. 19, 1997.

36. WARDLE, C.; DERAKHSHAN, H. Thinking about ‘information disorder’: formats of misinformation, disinformation, and mal-information. In: IRETON, C.; POSETTI, J. (org.). Journalism, ‘fake news’ & disinformation. Paris: UNESCO, 2018.

37. BAINES, D.; ELLIOTT, R. J.R. Defining misinformation, disinformation and malinformation: An urgent need for clarity during the COVID-19 infodemic. Discussion Papers, p. 20-06, 2020.

38. Bradshaw, S., Howard, P. N. The global organization of social media disinformation campaigns. Journal of International Affairs. 2018. Vol. 71. No. 1. P. 23-32.

39. Wardle, C., Derakhshan, H. Information disorder: Toward an interdisciplinary framework for research and policy making. Council of Europe. 2017. URL: https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-researc/168076277c (дата звернення: 20.05.2024).

40. Pennycook, G., Rand, D. G. Fighting misinformation on social media using crowdsourced judgments of news source quality. Proceedings of the National Academy of Sciences. 2019. Vol. 116. No. 7. P. 2521-2526.

41. Vosoughi, S., Roy, D., Aral, S. The spread of true and false news online. Science. 2018. Vol. 359. No. 6380. P. 1146-1151.

42. Allcott, H., Gentzkow, M. Social media and fake news in the 2016 election. Journal of Economic Perspectives. 2017. Vol. 31. No. 2. P. 211-236.

43. Conroy, N. K., Rubin, V. L., Chen, Y. Automatic deception detection: Methods for finding fake news. Proceedings of the Association for Information Science and Technology. 2015. Vol. 52. No. 1. P. 1-4.

44. Hasan, S., Qadir, J., Radev, D. R. The use of big data analytics in the fight against fake news. International Journal of Data Science and Analytics. 2016. Vol. 1. No. 1-2. P. 1-8.

45. Shao, C., Ciampaglia, G. L., Varol, O., Flammini, A., Menczer, F. The spread of fake news by social bots. Nature Communications. 2018. Vol. 9. No. 1. P. 4787.

46. Ferrara, E. The history of digital and social media analytics. The Journal of Technology in Society. 2020. Vol. 63. P. 101415.

47. Mihailidis, P., Viotty, S. Spreadable spectacle in digital culture: Civic expression, fake news, and the role of media literacies in “post-fact” society. American Behavioral Scientist. 2017. Vol. 61. No. 4. P. 441-454.

Додатки

ДОДАТОК А

Анкета для перевірки рівня медіаграмотності

1. Загальні відомості

1.1 Ваш вік:

- ☐ до 18
- ☐ 18-25
- ☐ 26-35
- ☐ 36-45
- ☐ 46-55
- ☐ 56 і більше

1.2 Стать:

- ☐ Чоловік
- ☐ Жінка
- ☐ Інше

1.3 Освіта:

- ☐ Середня
- ☐ Середня спеціальна
- ☐ Вища
- ☐ Інше (вкажіть) _____

2. Ваші знання про медіаграмотність

2.1 Як часто ви користуєтеся Інтернетом для отримання новин?

- ☐ Щодня

- ☐ Кілька разів на тиждень
- ☐ Раз на тиждень
- ☐ Рідше

2.2 Чи знаєте ви, що таке фейкові новини?

- ☐ Так
- ☐ Ні

2.3 Як би ви оцінили свою здатність розпізнавати фейкові новини?

- ☐ Дуже висока
- ☐ Висока
- ☐ Середня
- ☐ Низька
- ☐ Дуже низька

2.4 Які методи ви використовуєте для перевірки достовірності новин? (оберіть всі, що підходять)

- ☐ Перевіряю джерело новини
- ☐ Шукаю додаткову інформацію з інших джерел
- ☐ Читаю коментарі та відгуки інших користувачів
- ☐ Використовую спеціальні сервіси для перевірки фактів
- ☐ Інше (вкажіть) _____

3. Практичні навички

3.1 Як ви реагуєте на новини, які виглядають сумнівними?

- ☐ Ділюся ними в соціальних мережах
- ☐ Перевіряю достовірність перед поширенням

- ☐ Ігнорую їх
- ☐ Інше (вкажіть) _____

3.2 Чи брали ви участь у будь-яких програмах або заходах з підвищення медіаграмотності раніше?

- ☐ Так
- ☐ Ні

4. Перевірка знань

4.1 Оцініть, наскільки ви погоджуєтесь з наступними твердженнями:

Твердження	**Повністю згоден**	**Частково згоден**	**Не згоден**	**Важко сказати**
Я можу розпізнати фейкову новину.	[ ]	[ ]	[ ]	[ ]
Я знаю, як перевіряти достовірність інформації.	[ ]	[ ]	[ ]	[ ]
Я розумію, що таке медіаграмотність.	[ ]	[ ]	[ ]	[ ]
Я вмію користуватися інструментами для перевірки фактів.	[ ]	[ ]	[ ]	[ ]

5. Додаткові коментарі

5.1 Що, на вашу думку, найбільше допомогло б вам у підвищенні рівня медіаграмотності? _____

5.2 Ваші пропозиції або зауваження щодо освітніх програм з медіаграмотності: _____

ДОДАТОК Б

Тест на критичне мислення

Інструкція: Відповідайте на питання, обираючи одну або кілька відповідей, або надайте свою відповідь там, де це передбачено.

1. Аналіз аргументів

1.1 Прочитайте наступне твердження: "Більшість людей, які займаються спортом, ведуть здоровий спосіб життя. Тому, якщо ти хочеш бути здоровим, почни займатися спортом." Як ви оцінюєте цей аргумент?

- Дуже сильний
- Сильний
- Середній
- Слабкий
- Дуже слабкий

1.2 Які з наступних аргументів найбільш правильно підкріплюють твердження про користь спорту для здоров'я? (оберіть всі, що підходять)

- Наукові дослідження показують, що регулярні фізичні вправи покращують здоров'я серця.
- Мій друг почав займатися спортом і став почуватися краще.
- Всі спортсмени виглядають дуже здоровими.
- Лікарі рекомендують займатися спортом для підтримки фізичної форми.

2. Оцінка достовірності джерел

2.1 Яке з наступних джерел ви вважаєте найбільш надійним для отримання наукової інформації?

- Блог популярного фітнес-тренера
- Науковий журнал
- Соціальні мережі
- Вебсайт новин

2.2 Як ви перевіряєте достовірність інформації на незнайомому вебсайті? (оберіть всі, що підходять)

- Шукаю інформацію про автора
- Перевіряю дату публікації
- Переглядаю інші статті на вебсайті
- Порівнюю з іншими джерелами

3. Визначення упереджень

3.1 Прочитайте наступне твердження: "Цей бренд смартфонів є найкращим, оскільки його використовують багато відомих людей." Що може свідчити про упередженість цього твердження?

- Відсутність конкретних характеристик і переваг смартфонів
- Апеляція до популярності серед відомих людей
- Недостатність доказів або досліджень
- Усі відповіді вірні

3.2 Як ви оцінюєте об'єктивність статті, якщо вона містить багато емоційно забарвлених слів?

- Дуже об'єктивна
- Об'єктивна
- Частково об'єктивна
- Необ'єктивна

4. Логічні помилки

4.1 Визначте логічну помилку в наступному аргументі: "Якщо ми заборонимо всі ігри на мобільних телефонах, діти будуть краще вчитися."

- Хибна дилема
- Апеляція до страху
- Помилка композиції
- Причинно-наслідкова помилка

4.2 Прочитайте наступне твердження: "Всі мої знайомі кажуть, що цей фільм поганий, отже, він дійсно поганий." Яка логічна помилка тут присутня?

- Апеляція до авторитету
- Апеляція до популярності
- Хибна аналогія
- Апеляція до емоцій

5. Здатність до рефлексії

5.1 Як часто ви перевіряєте свої переконання та оцінюєте їх на основі нових даних?

- Дуже часто
- Часто
- Рідко
- Ніколи

5.2 Що ви робите, якщо знайдена вами інформація суперечить вашим попереднім переконанням?

- Ігнорую нову інформацію
- Перевіряю достовірність нової інформації
- Змінюю свої переконання
- Обговорюю з іншими для отримання додаткової думки

ДОДАТОК В

Приклад опитування для оцінки задоволеності користувачів заходами захисту від фейкових новин та їх впливу на поведінку користувачів у мережі:

Опитування задоволеності заходами захисту від фейкових новин

1. Загальна інформація

1.1 Ваш вік:

- до 18
- 18-25
- 26-35
- 36-45
- 46-55
- 56 і більше

1.2 Стать:

- Чоловік
- Жінка
- Інше

1.3 Ваш рівень освіти:

- Середня
- Середня спеціальна
- Вища
- Інше (вкажіть) _____

2. Використання інструментів захисту від фейкових новин

2.1 Які інструменти захисту від фейкових новин ви використовуєте? (оберіть всі, що підходять)

- Антивірусне програмне забезпечення

- Браузерні розширення для перевірки фактів
- Спеціалізовані вебсайти для перевірки фактів
- Соціальні мережі з вбудованими інструментами перевірки
- Інше (вкажіть) _____

2.2 Як часто ви користуєтеся цими інструментами?

- Щодня
- Кілька разів на тиждень
- Раз на тиждень
- Рідше

3. Оцінка зручності використання інструментів

3.1 Наскільки зручно вам використовувати інструменти захисту від фейкових новин?

- Дуже зручно
- Зручно
- Нейтрально
- Незручно
- Дуже незручно

3.2 Які з наступних факторів найбільше впливають на зручність використання інструментів? (оберіть всі, що підходять)

- Інтерфейс користувача
- Швидкість роботи
- Точність результатів
- Доступність і легкість налаштування
- Інше (вкажіть) _____

4. Оцінка ефективності інструментів

4.1 Як би ви оцінили ефективність інструментів у захисті від фейкових новин?

- Дуже ефективні
- Ефективні
- Нейтральні
- Неефективні
- Дуже неефективні

4.2 Чи зменшилася кількість фейкових новин, які ви помічаєте, з початку використання цих інструментів?

- Значно зменшилася
- Зменшилася
- Не змінилася
- Збільшилася
- Значно збільшилася

5. Вплив на поведінку в мережі

5.1 Як використання інструментів захисту від фейкових новин вплинуло на вашу поведінку в мережі?

- Став більш уважним до джерел інформації
- Почав перевіряти інформацію перед тим, як поділитися нею
- Зменшив використання соціальних мереж
- Не помітив змін
- Інше (вкажіть) _____

5.2 Чи змінилося ваше ставлення до новин, які ви читаєте в Інтернеті, після використання інструментів захисту?

- Так, я став більш скептичним
- Так, я став більше довіряти перевіреним джерелам
- Ні, моє ставлення не змінилося
- Інше (вкажіть) _____

6. Додаткові коментарі та пропозиції

6.1 Що, на вашу думку, можна покращити в інструментах захисту від фейкових новин? _____

6.2 Ваші загальні враження та пропозиції щодо заходів захисту від фейкових новин: _____