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: “МЕТОДИ РОЗПІЗНАВАННЯ ФЕЙКОВИХ ПРОФІЛІВ В ПРОФЕСІЙНИХ 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

Нікіта ГУРІНОВ
(підпис) Ім’я, ПРІЗВИЩЕ здобувача

Виконав: здобувач вищої освіти гр. УБД-41

Нікіта ГУРІНОВ
Ім’я, ПРІЗВИЩЕ

Керівник: Віталій ТИЩЕНКО
Ім’я, ПРІЗВИЩЕ

Рецензент: _____
К.т.н., доцент Ім’я, ПРІЗВИЩЕ

Київ 2025

ДЕРЖАВНИЙ УНІВЕРСИТЕТ
ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ

Навчально-науковий інститут кібербезпеки та захисту інформації

Кафедра управління кібербезпекою та захистом інформації

Ступінь вищої освіти бакалавр

Спеціальність 125 Кібербезпека

Освітня програма Управління інформаційною та кібернетичною безпекою

ЗАТВЕРДЖУЮ
Завідувач кафедри УКБЗІ
_____ Світлана ЛЕГОМІНОВА
“____”_____2025 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

Гурінову Нікіті Володимировичу
(прізвище, ім’я, по батькові здобувача)

1. Тема кваліфікаційної роботи “Методи розпізнавання фейкових профілів в професійних соціальних мережах”,
керівник кваліфікаційної роботи ТИЩЕНКО Віталій.
(ІПРІЗВИЩЕ, Ім’я,, науковий ступінь, вчене звання)

затверджені наказом Державного університету інформаційно-комунікаційних технологій "Про закріплення тем випускних кваліфікаційних робіт та призначення наукових керівників на 2024-2025 н.р. за студентами першого (бакалаврського) рівня вищої освіти". № 195 від 07.03.25

2. Строк подання кваліфікаційної роботи “25” травня 2025р.
3. Вихідні дані до кваліфікаційної роботи: *методи машинного навчання, інструменти аналізу профілів користувачів соціальних мереж, методи та засоби забезпечення інформаційної безпеки, міжнародні стандарти, наукова та технічна література*.
4. Перелік питань, які мають бути розроблені:
   4.1. Проаналізувати поняття фейків у текстовому контенті.
   4.2. Дослідити методи виявлення фейкових профілів користувачів соціальних мереж.
   4.3. Вивчити процес розробки системи машинного навчання.
5. Перелік ілюстративного матеріалу: *презентація PowerPoint*

6. Дата видачі завдання “5” березня 2025 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Етапи кваліфікаційної роботи	Термін виконання етапів роботи	Примітка
1.	Визначення об’єкту, предмету, мети та завдань дослідження.	18.03.2025	
2.	Збір та аналіз літератури.	29.03.2025	
3.	Аналіз поняття фейків у текстовому контенті.	08.04.2025	
4.	Дослідження методів виявлення фейкових профілів користувачів соціальних мереж.	22.04.2025	
5.	Вивчення процесу розробки системи машинного навчання.	08.05.2025	
6.	Формулювання висновків за результатами проведеного дослідження.	20.05.2025	
7.	Оформлення роботи.	22.05.2025	
8.	Оформлення презентації.	03.06.2025	
9.	Отримання рецензії на роботу.	03.06.2025	
10.	Захист в ЕК.	__.06.2025	

Здобувач вищої освіти _____ Нікіта ГУРІНОВ
(підпис) *(Ім’я, ПРІЗВИЩЕ)*

Керівник кваліфікаційної роботи _____ Віталій ТИЩЕНКО
(підпис) *(Ім’я, ПРІЗВИЩЕ)*

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

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

Направляється здобувач Гурінов Н.В. до захисту кваліфікаційної роботи
(прізвище та ініціали)
за спеціальністю 125 Кібербезпека
(код, найменування спеціальності)
освітньої програми Управління інформаційною та кібернетичною безпекою
(назва)
на тему: "Методи розпізнавання фейкових профілів в професійних соціальних мережах"

Кваліфікаційна робота і рецензія додаються.

Директор ННІКБЗІ _____ Євгенія ІВАНЧЕНКО
(підпис) *(Ім'я, ПРІЗВИЩЕ)*

Висновок керівника кваліфікаційної роботи

Здобувач у кваліфікаційній роботі проаналізував особливості розробки системи машинного навчання для виявлення фейків у текстовому контенті, дослідив основні методи та технології обробки природної мови, вивчив алгоритми машинного навчання для класифікації тексту, розробив практичні рекомендації за темою дослідження.

Автор показав розуміння проблеми дослідження та бачення основних теоретичних і практичних напрямів її вирішення, довів володіння методами наукового дослідження, проявив себе як організований, відповідальний виконавець. Результати дослідження апробовані на двох конференціях.

Все це дозволяє оцінити кваліфікаційну роботу здобувача ГУРІНОВА Нікіти на оцінку "_____" та присвоїти йому кваліфікацію бакалавра з кібербезпеки за освітньою програмою Управління інформаційною та кібернетичною безпекою.

Керівник кваліфікаційної роботи _____ Віталій ТИЩЕНКО
(підпис) *(Ім'я, ПРІЗВИЩЕ)*

"____"_____2025 року

Висновок кафедри про кваліфікаційну роботу

Кваліфікаційна робота розглянута. Здобувач Гурінов Н.В. допускається до захисту даної роботи в Екзаменаційній комісії.

Завідувач кафедри управління кібербезпекою та захистом інформації _____ Світлана ЛЕГОМІНОВА
(підпис) *(Ім'я, ПРІЗВИЩЕ)*

ВІДГУК РЕЦЕНЗЕНТА
на кваліфікаційну бакалаврську роботу

здобувача вищої освіти ГУРІНОВА Нікіти
на тему “Методи розпізнавання фейкових профілів в професійних соціальних мережах”

Актуальність. У сучасному світі, де зростає залежність суспільства від інформаційних технологій та кіберпростору, тема розробки системи машинного навчання для виявлення фейків у текстовому контенті є надзвичайно актуальною. Поширення фейкових новин та дезінформації може мати серйозні наслідки для громадської думки, політичної стабільності та економічного благополуччя. Тому розробка ефективних методів для автоматичного виявлення фейкових новин є критично важливою для забезпечення достовірності інформації та захисту
суспільства від маніпуляцій.

Позитивні сторони.

1. В роботі використано сучасні методи машинного навчання, що дозволяють ефективно виявляти фейкові новини. Це свідчить про високий рівень теоретичної підготовки здобувача.
2. Розроблена система має значний потенціал для практичного застосування, зокрема, у засобах масової інформації та соціальних мережах для автоматичного фільтрування контенту.
3. Здобувач провів всебічний аналіз існуючих методів і підходів до виявлення фейків, що дозволило вибрати оптимальне рішення для розробки власної системи.
4. Прототип системи показав високу ефективність на тестових даних, що підтверджує правильність обраних методів та підходів.
5. Робота відзначається високою якістю оформлення, логічною структурою та чіткістю викладу матеріалу.

Недоліки.

1. Тестування системи проводилося на відносно невеликому наборі даних, що може не повністю відображати її ефективність в реальних умовах.
2. Існує потенціал для подальшого вдосконалення алгоритмів з метою підвищення

Висновок: Кваліфікаційна робота виконана на належному науково-методичному рівні і заслуговує оцінки “_____”, а здобувач заслуговує присвоєння кваліфікації бакалавра з кібербезпеки за
освітньою програмою Управління інформаційною та кібернетичною безпекою.

Рецензент: _____ _____
підпис Ім’я,
ПРІЗВИЩЕ

РЕФЕРАТ

Кваліфікаційна робота присвячена розробці автоматизованої системи машинного навчання для виявлення фейкових профілів користувачів соціальних мереж. Робота складається зі вступу, трьох розділів, що містять 12 рисунків, висновків і списку використаних джерел із 74 найменувань. Загальний обсяг роботи становить 83 аркуші, з яких 9 аркушів займають перелік умовних скорочень та список використаних джерел.

Метою кваліфікаційної роботи є покращення забезпечення кібербезпеки шляхом розробки метрик та системи підтримки прийняття рішень для визначення фейкових облікових записів у соціальних мережах.

Об'єкт дослідження – профілі користувачів у соціальних мережах.

Предмет дослідження **-** методи розпізнавання фейкових профілів користувачів.

Методи дослідження. Для вирішення означеного наукового завдання в роботі використані методи аналізу та синтезу, машинного навчання, обробки природної мови (NLP), порівняння, класифікації, та експертної оцінки.

У роботі розглянуто та проаналізовано структуру облікових записів користувачів у соціальних мережах. Запропоновано метрики ознак фейкових облікових записів у категоріях «Лайки», «Друзі», «Пости та статуси», «Персональна інформація» та «Фото», а також їх можливі параметри та ступінь впливу на фейковість облікового запису. Розроблені структурні моделі, що дозволяють виявляти фейкові облікові записи у соціальних мережах. Проаналізовано існуючі і розроблено власну систему підтримки прийняття рішень для аналізу профілів користувачів у соціальній мережі «Facebook». Мова програмування - Python. Проведено низку експериментальних досліджень для перевірки роботи програмного засобу шляхом аналізу облікових записів у соціальній мережі «Facebook».

Галузь застосування. Розроблені підходи можуть бути використані для автоматичного виявлення фейкових профілів користувачів у соціальних

мережах та інших інформаційних ресурсах.

Ключові слова: фейкові профілі користувачів, системи прийняття рішень.

ABSTRACT

The qualification work is devoted to the development of an automated machine learning system for detecting fake profiles of social network users. The work consists of an introduction, three sections containing 12 figures, conclusions and a list of sources used from 74 names. The total volume of the work is 83 sheets, of which 9 sheets are occupied by a list of conventional abbreviations and a list of sources used.

The purpose of the qualification work is to improve cybersecurity by developing metrics and a decision support system for identifying fake accounts in social networks.

The object of the study is user profiles in social networks.

The subject of the study is methods for recognizing fake user profiles.

Research methods. To solve the specified scientific problem, the work used methods of analysis and synthesis, machine learning, natural language processing (NLP), comparison, classification, and expert assessment.

The work considers and analyzes the structure of user accounts in social networks. Metrics of signs of fake accounts in the categories "Likes", "Friends", "Posts and statuses", "Personal information" and "Photos" are proposed, as well as їх possible parameters and the degree of influence on the fakeness of the account. Structural models have been developed that allow detecting fake accounts in social networks. Existing ones have been analyzed and our own decision support system for analyzing user profiles in the social network "Facebook" has been developed. Programming language - Python. A number of experimental studies have been conducted to verify the operation of the software tool by analyzing accounts in the social network "Facebook".

Field of application. The developed approaches can be used to automatically detect fake user profiles in social networks and other information resources.

Keywords: fake user profiles, decision-making systems.

ЗМІСТ

ВСТУП

В часи постійного вдосконалення інформаційних та комунікативних технологій, будь-який конфлікт має своє відображення в мережі Інтернет. Дуже часто такі повідомлення впливають на результати протиборства конкуруючих сторін. Активне залучення до віртуальної мережі багатомільйонної аудиторії дозволяє маніпулювати громадською думкою і суттєво впливати на середовище [1].

Розуміння значення Інтернету та соціальних мереж змушує передові держави вкладати величезні інвестиції в соціальні мережі, які нині стають не тільки засобом комунікації, а й ефективним інструментом політичного впливу. Сьогодні соціальні мережі використовуються у якості платформи для проведення спеціалізованих інформаційних дій, таких як інформаційно-психологічні операції, спрямовані на формування думки суспільства [2, 3]. Величезна кількість людей зі всього світу щодня використовують соціальні мережі як засіб для спілкування, джерело інформаційних новин, перегляд розважального контенту, проте є також інша група людей, які за допомогою інформаційних “вкидів” маніпулюють індивідуальною та суспільною свідомістю [2]. Для цього зазвичай використовуються фейкові облікові записи, на яких відсутня інформація про користувача або розміщені неправдиві дані. Часто фейкові облікові записи використовуються для прямої чи опосередкованої зміни думки соціуму у різних формах прояву. Оскільки у соціальних мережах немає жорсткої цензури, а іноді вона і зовсім відсутня, це сприяє успішності таких дій у віртуальному світі. Таким чином, наведені факти вказують на те, що соціальні мережі перетворюються на своєрідний майданчик для інформаційних баталій.

Проте, формалізованих моделей, на базі яких можна розробити інструментарій для аналізу облікових записів щодо їх автентичності, не існує, тому будь які дослідження, виконані в цьому напрямку, є актуальними.

Актуальність теми. У сучасному світі, де зростає залежність

суспільства від інформаційних технологій та кіберпростору, тема розробки системи машинного навчання для виявлення фейків є надзвичайно актуальною. Поширення фейкових новин та дезінформації може мати серйозні наслідки для громадської думки, політичної стабільності та економічного благополуччя. Тому розробка ефективних методів для автоматичного виявлення фейків є критично важливою для забезпечення достовірності інформації, що циркулює, та захисту суспільства від можливих маніпуляцій.

Метою кваліфікаційної роботи є покращення забезпечення кібербезпеки шляхом розробки метрик та системи підтримки прийняття рішень для визначення фейкових облікових записів у соціальних мережах.

Об'єкт дослідження – профілі користувачів у соціальних мережах.

Предмет дослідження **-** методи розпізнавання фейкових профілів користувачів соціальних мереж.

Для досягнення мети необхідно виконати такі **завдання**:

1. Проаналізувати існуючі підходи до аналізу облікових записів користувачів соціальних мереж.
2. Розробити структурні моделі ознак фейкових облікових записів.
3. Розробити архітектуру системи прийняття рішень для розпізнавання фейкових облікових записів у соціальних мережах.
4. Розробити програмний засіб для автоматизації виявлення фейкових облікових записів у соціальних мережах.
5. Провести експериментальні дослідження для тестування роботи програми.

Методи дослідження**.** Для вирішення означеного наукового завдання в роботі використані методи аналізу та синтезу, машинного навчання, обробки природної мови (NLP), порівняння, класифікації, та експертної оцінки.

Як результат, у роботі досліджено технології обробки природної мови для виявлення фейкових новин, проаналізовано відомі методи машинного навчання, розроблено та протестовано систему машинного навчання, запропоновано практичні рекомендації щодо її застосування для виявлення

фейків.

Наукова новизна полягає у вдосконаленні методу виявлення фейкових облікових записів у соціальних мережах шляхом розробки метрик ознак фейковості, що дозволяє збільшити достовірність прийняття рішення до 94 %.

Практичне значення одержаних результатів. полягає у розробці алгоритмів та програмного забезпечення виявлення фейкових облікових записів у соціальній мережі «Facebook», що дозволяє виявляти інформаційних агентів-ботів під час інформаційної війни.

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

РОЗДІЛ 1. ПОНЯТТЯ ФЕЙКІВ, ЇХ КЛАСИФІКАЦІЯ, МЕТОДИ РОЗПІЗНАВАННЯ

1.2 Поняття фейків у цифровому середовищі

Власне, слово **фейк** з'явилося у щоденному вжитку досить нещодавно, перші приклади його використання у мовленні відносяться до 70-х років XX ст., однак самі фейки зустрічалися у житті людей набагато раніше, проте не асоціювалися із цим словом.

У перекладі з англійської мови слово "**фейк**" означає *підробка* або *фальшивка*. У контексті нашого дослідження – це фальсифікація певних відомостей з ціллю омани людей. Люди ще у середньовіччі вдавалися до фальсифікації документів, додавання до них певних коректив, прагнучи отримати якість моральні чи фінансові вигоди для себе. Крім того, фейки активно використовувалися для пропаганди, наприклад, під час проведення виборів (керівника), для формування авторитету певної особи чи режиму, руйнування впливу суперників тощо.

Яскравим прикладом в історії України можна навести діяльність КДБ СРСР щодо українського медіа "Радіо Свобода", щодо якого велася активна пропагандистська діяльність з метою ліквідації мовлення та переконання слухачів у недостовірності інформації, що надійшла з цього джерела. КДБ множило хибні інформаційні повідомлення стосовно працівників даного радіо, звинувачувало інші країни у розвідницькій та пропагандистській діяльності через це джерело на їх користь. «Більшість американців, що працюють на станції, є євреями за національністю, вихідцями з території СРСР. Усі керівні посади зайняті, як правило, американцями, значний відсоток з яких складають співробітники ЦРУ» -вказано у документах КДБ за 1969 р. [1].

Часто це були необґрунтовані звинувачення, не підкріплені доказами. Повідомлення такого типу традиційно були утворені як яскраве ствердження окремого факту. Оскільки це джерело об'єднувало в собі багато підрозділів з

різних держав, то для певної достатньо великої частини населення такі заяви видавалися достовірними. Також КДБ намагався спотворювати факти, які стосувалися органів влади, подані на “Радіо Свобода”. Це безсумнівно можна віднести до категорії фейків. Сюди ж відносяться плакати з текстами, які звинувачують “Радіо Свобода” у нацизмі й співпраці з Заходом. Підкреслимо, що радянський лад завжди будувався на пропаганді й спотворенні правди. Відтак, він патологічно боявся висвітлення правди у медіа. Тому постійно намагався підроблювати факти або забороняти вільне поширення інформації. Людям подавалася викривлена реальність, прикрашена різноманітними пропагандистськими наративами.

Справді, неправдиву інформацію досить легко замаскувати під гучними заголовками та грамотним поданням матеріалу, що здатний зацікавити людей. В якості гарної ілюстрації можна навести випадок, втілений у життя американський публіцистом С. Глассом. Протягом трьох років роботи у журналі він зумів сфабрикувати більш як половину власних статей. Зауважимо, що протягом цього відрізку часу його популярність стрімко зростала. Він навіть встиг виготовити сайт компанії, стосовно якої готував публікацію, щоб підтвердити її правдивість. Проте саме на ній і було викрито його підробку.

Таким чином, можна зробити висновки, що читачі рідко перевіряють та аналізують отриману інформацію. Навіть у випадку друку статті у виданні, якому ми довіряємо, інформація може виявитися фейком. В результаті це може суттєво вплинути на наше світосприйняття і наше життя. Тож фейк насправді є дуже небезпечною зброєю. Її потрібно вміти вчасно ідентифікувати для усунення небезпеки.

1.2 Класифікація фейків

Незамінним кроком, що передує розробці способів розпізнавання фейків є розуміння їх класифікації. Саме вона здатна сформувати чітко визначене розуміння видів фейків і їх властивостей. Найбільш вагома цінність

класифікації фейків міститься у наступному: по-перше, вона дозволить розпізнавати і об'єднувати фейки за певними властивостями, такими як *спосіб їх продукування, типи маніпуляцій, використані медіа формати* тощо. По-друге, маючи у розпорядженні коректну класифікацію, ми будемо зосереджувати зусилля якраз на формуванні адекватних методів розпізнавання для фейків окремих типів. По-третє, це сприятиме встановленню основних ділянок, де вони переважно поширюються або де здійснюють найбільш вагомий вплив. Це буде актуальним у процесі розробки пріоритетів і плануванні ресурсів для протидії фейкам. Крім того, вона дозволить встановити базові закономірності й характерні відмінності, що притаманні певному типу фейків. Це дозволить описати спеціалізовані моделі й розробити алгоритми, що здатні ефективно виявляти фейки різних типів.

Таким чином, класифікація фейків є необхідним етапом у організації протидії появі фейків. Вона дозволяє охопити наявну множину фейків і розробити практичні методи їх розпізнавання й протидії поширенню.

1.2.1 Спосіб класифікації фейків, запропонований в університеті Онтаріо

Однієї єдиної класифікації фейків не існує. Різноманітні наукові організації і засоби медіа постійно намагалися запропонувати хоча б один їх стандартизований опис. Вченими з університету Онтаріо свого часу було запропоновано поділяти фейки на три категорії:

1. **Серйозні фабрикації** (фейки, які було заплановано й створено заздалегідь). Сюди можна відносити шахрайські репортажі, спрямовані на обдурювання й захоплення уваги, жовта преса, що продукує гучні заголовки й цікаві теми, що захоплюють аудиторію і за рахунок цього забезпечують збільшення прибутків таблоїда внаслідок збільшення його популярності. Найбільш часто вживаними темами, які спостерігаємо у жовтій пресі, є випадки, тісно пов'язані з криміналом (наприклад, вбивства, грабежі тощо),

історії з подробицями життя зірок чи інших відомих людей, політиків тощо, різноманітні сенсаційні випадки.

2. **Масштабні містифікації**: цього стосуються різні вигадки, поширювані у соціальних мережах. Наприклад, інформація розміщена на сторінці популярного блогу, може бути створена схожою на новину з метою введення в обману широкого загалу.

3. **Гумористичні фейки**: гумор, жарти, сатира, до якої може бути докладено певну інформацію про реальні події. Традиційно люди не ставляться серйозно до інформації з подібних джерел, адже їх мета – розважальна. Проте у будь-якому жарті зазвичай вміщується краплина правди, відтак завжди буде існувати така категорія людей, які сприймуть жарт за чисту правду.

В якості прикладу наведемо **The Onion** – американську газету іронічних новин, а також відповідні веб сайти, які вміщують різнопланові жартівливі новини. Агенція дуже популярна, відтак вона здатна виступити в якості загрози, адже по суті вона вся начинена фейками, що за допомогою аудиторії безупинно поширюються і тим самим уводять в оману великі маси людей.

1.2.2 Класифікація фейкових новин запропонована BBC

Трохи інакшу класифікацію фейків пропонує BBC:

- Спочатку вони виокремлюють **спеціально сфальсифіковані новини**, їх поширення здійснюється зумисно, тобто їх розміщують у популярних журналах чи інтернет виданнях (медіа) зумисно, з ініціативи конкретної людини чи групи людей, які завдяки цим діям планують отримати особисту матеріальну, моральну чи будь-яку іншу вигоду.
- Іншим видом є **неточні новини**, тобто такі, що містять у собі більшу частину правдивих даних, але певний відсоток новини може бути перебільшено, або викривлено, або зовсім неправдиво подано. Це може бути факт, дата, інформація про учасників тощо.

1.3. Дослідження сучасних методів виявлення фейків

Нині відомо багато різних підходів до розпізнавання фейкової інформації, особливо фейкових фотографій, рисунків, які дуже часто розміщуються у різних соціальних мережах та веб ресурсах. Часом це відбувається через дефіцит реальних зображень для супроводу певної новини. Присутність у репортажі фото чи відео, постерів розширює спектр емоцій, що виникають у читачів під час перегляду інформації та суттєво змінює її сприйняття.

Іншим варіантом використання є такий, коли старі фото чи відео частково змінюються і видаються за нові з метою роздування описаних подій. Це часто трапляється з фотографіями з місць відомих катастроф, катаклізмів, воєн тощо. Зазвичай такі фото можна легко доповнити або переробити для видання однієї ситуації за іншу і це видається авторам публікацій набагато простішим, ніж збирати фото з реального місця події. У якості прикладу можна навести фото війни з Сирії, змінені певними виданнями та розміщені в Інтернеті як фото-звіти про бойові дії на сході України 2014 року.

На даний час видання старих фото за нові є поширеною проблемою через нестабільний стан у світі та емоційну складову в цьому, тож допомога звичайним громадянам у виявленні фальсифікованої інформації є актуальною проблемою. Нині вже існують спеціальні програмні сервіси, що здійснюють пошук за фото. Найпростіші варіанти програм такого типу - це плагіни, які встановлюються у пошуковий браузер. За допомогою подібного плагіну при обранні фото (відео) ми можемо знайти джерела і дати публікацій, у яких воно використовувалося раніше і, основуючись на цьому, зробити певні висновки. Прикладом такого плагіну є **RevEye**, який використовується для **Google Chrome** і функціонує за зазначеним раніше сценарієм. Ще одним прикладом, який вже можна використовувати у декількох браузерах, є плагін Who stole my pictures. Він розроблювався для перевірки готових публікацій на їх поширення іншими користувачами або ресурсами. В основному плагін має той самий принцип дії, що і попередній плагін для **Google Chrome**. Крім того, існують

окремі програми, що вчиняють пошук фото за геолокацією та датою його створення. Це допомагає у випадках проведення слідчих експериментів, пошуку свідків певної події та загалом огляду зазначеної локації з метою отримання додаткової інформації.

Проаналізувавши низку таких програм, можна стверджувати, що всі вони реалізують досить різноманітні підходи з метою перевірки правдивості чи неправдивості інформації. Далі розглянемо, які саме складові є важливими для ідентифікації фейків та які взагалі існують моменти, на які варто звертати увагу у процесі їх аналізу.

1.3.1. Візуальна складова у розпізнаванні фейків

Графічна частина довільної публікації є одним з основних моментів, яким ми приділяємо увагу в процесі аналізу даних. Фото вміщує важливі моменти стосовно власне інформації, а також про її належність до фейків. Пошуковці галузі ідентифікації фейків за графічними включеннями поділили фейки за окремими властивостями: криміналістичні ознаки, семантика, зміст і статистика.

Під **криміналістичними** маємо на увазі пошук ознак підробки фото або відео, які творець залишив після себе (відомо, що скоригувати щось безслідно практично неможливо). Це можуть бути інтервали, що залишилися в процесі поєднання різних відео, зміна якості зображення у окремих фрагментах відео, гама кольорів, інтерполяція, шуми тощо. Для виявлення таких дій зазвичай використовують бінарні патерни на зображенні, які змінюються відповідно до проведеної над ними маніпуляції. Широке застосування нейронних мереж також додало роботи слідчим, які спеціалізуються по фейковим новинам, адже штучний інтелект активно використовується для генерування нового медіа з високою долею правдоподібності. Загалом слідча експертиза у таких випадках базується на аналізі сигналів, спектральних характеристиках та кореляції, але оскільки більшість підробок базується на підробленні відео контенту з участю

людей то припущення щодо фейковості робилися і на основі візуальної складової, а саме - на недостатньо природній поведінці персонажу у кадрі, деформації його обличчя тощо.

1.3.2. Роль семантики у процесі розпізнавання фейкової інформації

Не менш важливою ознакою, яка допомагає відрізняти фейкову новину від звичайної, є семантична складова. Як вже було вказано раніше, автори фейкових новин зазвичай подають їх у вигляді, що потенційно викликатиме якнайбільше переживань у читачів, тому графічний супровід неймовірно важливий, адже саме він керує емоціями людини. Найбільш потужним засобом вивчення семантики візуального контенту є згорткова нейронна мережа VGG від CNN. Вона вживається для графіки різних типів. Крім того, її використання можливе для дослідження фейків. Ця нейронна мережа будується з 13 згорткових шарів і 3 основних. Зосередимося далі на її архітектурі:

- Вхід, на який отримуємо набір пікселів зазначеного розміру, а саме 244×244.
- Згорткові шари, що виокремлюють головні ознаки рисунка, і всі трансформації, які було накладено, для наступного аналізу.
- Об’єднання шарів спрямоване на спрощення аналізу за рахунок зменшення кількості ознак для нього.
- З’єднані шари націлені на класифікацію зображень.

Окрім цієї у CNN є інші моделі, проте вони переважно спрямовані на пошук фейкових новин, які функціонують на основі об’єднання інформації, що базується на аналізі пікселів та частот. Також є моделі для пошуку певних емоційних провокацій, що включаються до фейкової інфомації і є складовою семантичної частини.

1.3.3. Використання статистичних ознак для ідентифікації фейкових повідомлень

Наступною ознакою для аналізу фейкових новин є **статистичні риси**. Інакше кажучи, **статистична риса** відповідає за кількість зображень, що відповідають тій чи іншій новині, та їх мінливість. На нашу думку, має видаватися досить підозрілим, коли при гортанні стрічки новин ми на певну екстраординарну подію бачимо тільки одне фото. Навряд чи журналісти, які здійснювали репортаж з місця подій, зробили недостатньо фотографій для підтвердження висвітлюваної новини, або взагалі на незалежних ресурсах ми бачимо одні й ті ж фото, зняті з однакового ракурсу. Тому доцільно проводити певні обчислення, а саме - можна ввести статистичні функції, що оброблюють аналіз розподілу правдивих новин і фейків. Розглянемо основні характеристики, на яких базуються роботи з проведення статистичного аналізу:

- **Кількісні відношення**: прикладом цього при аналізі новини є зіставлення кількості зображень, використаних для її опису, з загальною кількістю застосунків, що стосуються висвітлення подій, описаних у повідомленнях.
- **Відношення популярності**: аналіз основується на популярних рисунках, взятих зі стрічки, які можуть мати багато відміток, вподобайок, збережень, перенадсилань тощо, і співставленні їх з типовими дописами.
- **Аналіз типів зображень**: читаючи різні новини, ми зустрічаємо зображення різних форматів, іноді те чи інше джерело намагається слідкувати за збереженням єдиного формату, але іноді ми зустрічаємо зовсім різні, тож аналізуючи й статистично виводячи найбільш популярний формат серед потоку, ми можемо далі аналізувати інші зображення, основуючись на ньому.

Ми описали базові статистичні прийоми, проте можна згадати і про глибші (наукові):

- **Оцінка чіткості**: метод Куллбака-Лейблера аналізує розбіжності між певними сетами зображень – колекцією (збірка зображень з певних подій) та подією (фото з події, яка аналізується). Аналіз проводиться за формулою:

$$VCS = DKL(p(w|c)||p(w|k)), \quad (1.1)$$

де $p(w|c)$ та $p(w|k)$ оцінюють частоту появи зображення певного слова w, у колекції або у самій події відповідно.

- **Оцінка когерентності**: функція подібності використовується для порівняння зображень з колекції події, яка оцінюється, на схожість (когерентність).
- **Оцінка через гістограму розподілу**: для оцінки події цільової новини створюють матрицю подібності зображень, з неї отримують певний вектор подібності зображень і будують гістограму розподілу подібності зображень на рівні дрібної деталізації.
- **Оцінка багатогранності візуальної складової**: оцінюється певна візуальна відмінність між окремими екземплярами колекції, в якості бази обирають найбільш популярне зображення або таке, що краще відповідає новині, або краще за якістю; на його основі проводиться ранжування набору інших зображень з цієї колекції.
- **Оцінка через кластеризацію**: метод кластеризації використовується для оцінки набору зображень стосовно певної новини, для цього використовується алгоритм ієрархічної агломеративної кластеризації.

1.3.4. Аналіз інформації з урахуванням контексту та метаданих

Обов'язковою складовою підтвердження фейковості новини є її **контекст**. Зауважимо, що перевірки на відповідність тексту зображенням не забезпечують очікуваного результату, адже автори фейкових новин зазвичай узгоджують підписи з інформацією у власних дописах. Тому цей метод краще підходить для оцінки інших прихованих факторів, таких як відповідність наповнення веб сторінки, на якій опубліковано новину, її візуальному вмісту і метаданим.

По-перше, розглянемо **метадані**. Сюди віносяться дані про медіа файли, дату, час та місце їх створення, розмір, розширення тощо. Маючи їх у

розпорядженні, ми можемо використати всю необхідну нам інформацію для подальшої оцінки. Ці дані дуже часто губляться після публікації і тому надалі вже не забезпечують нам бажаний результат. Проте у випадку наявності вони можуть стати надійним доведенням фейковості.

По-друге, не менш активно працюють характеристики контексту, отримані зовні в результаті зворотнього пошуку фото. Зворотній пошук зображень виокремлюється від традиційного тим, що він отримує зображення в якості вхідних даних і у відповідь повертає перелік релевантних веб-сторінок, що вміщують відповідне зображення, супроводжуючи його назвою, описом та часом. Процес досить легко автоматизувати, застосувавши до величезної кількості зображень за допомогою API пошукових систем, таких як зворотній пошук зображень Google.

Далі деталізуємо наступні **функції контексту**:

- **Часові проміжки**: визначаються як проміжки у часі між моментом публікації новини та моментом першої публікації візуального вмісту. Ця функція може бути використана для перевірки оригінальності візуального контенту. Якщо часовий інтервал перевищує певне порогове значення, то ймовірність того, що візуальний вміст походить від нерелевантної події, збільшується.
- **Схожість між новинами**: подібність між новинами визначається як схожість між заявою у новині та текстовим вмістом сканованих веб-сайтів. Беремо до уваги наступне: повідомлення веб-сайтів є корисними для тлумачення автентичності фотографій з події. Функція стосується перевірки узгодженості між текстовим супроводом та графічним контентом, який пропонується для цієї події.
- **Надійність платформи**: соди ми вкладаємо ставлення до платформи, де міститься набір фотографій з події. Основуючись на вживанні даних з веб сайту, що містить достовірні дані про більш ніж 2700 медіа-джерел, усі веб-сторінки, повернуті методом зворотного пошуку, було класифіковано на такі групи: висока, низька чи змішана фактичність. Частка сторінок з кожної

категорії була поставлена у відповідність поняттю надійності платформи.

1.3.5. Вплив зображень та графіки на процес аналізу новин

Вище нами наведено такі види функцій класифікації фейкових новин: **криміналістичні, семантичні, статистичні** та **контекстні**. Вони виявляють окремі властивості графічного наповнення, тому їх доволі часто комбінують для покриття більшої множини сценаріїв фальсифікації. Не менш важливо дослідити способи ідентифікації фейкових новин, що орієнтуються на візуальний контент. Їх розділяють на такі, що базуються на **контенті** або на **знаннях**.

Способи засновані на контенті передбачають пошук і суміщення сигналів різноманітних модальностей із контенту. Вони не орієнтуються на наявні дані.

Способи засновані на знаннях передбачають перевірку фактів на наявних даних, орієнтуючись на зовнішні джерела. Передбачається наявність досить потужного сету еталонних даних. Оцінка новини виконується шляхом порівняння з наявними схожими за змістом дописами, що містяться у базі даних.

Будь-яка новина поєднує текст і фотографії. Обидві частини можуть забезпечити доказами, на базі яких можна доводити фейковість. Сучасні пошуки за цією тематикою були націлені на моделювання поєднання інформації з різних модальностей. В першу чергу вони орієнтувалися на рекурентну нейронну мережу (RNN), попередньо навчену CNN для отримання окремих семантичних характеристик.

Новітні пошуки, спрямовані на виявлення фейкових новин, поєднують інформацію, отриману з кількох характерних зрізів. Інакше кажучи, вони оперують з мультимодальними даними. Вперше мультимодальний контент було включено через глибокі нейронні мережі. Було запропоновано інноваційний RNN для поєднання функцій текстового, графічного і соціального наповнення. Першочергово текст і соціальний зміст поєднуються з LSTM для

єдиного подання. Зрештою, потім воно змішується з візуальними характеристиками, які витягують із заздалегідь навченого глибокого CNN.

На кожному проміжку часу вихідні дані LSTM використовуються на рівні нейронів для узгодження окремих характеристик під час об’єднування.

Була запропонована нейронна мережа подій на основі мультимодальних функцій, інваріантних до однієї події. Мережа вміщує такі компоненти:

- мультимодальний екстрактор,
- детектор фейкових новин,
- дискримінатор подій.

Мультимодальний екстрактор спрямований на витягування текстів і графіки з новин. Екстрактор стикується з детектором, щоб навчитися дискримінаційному поданню для детекції фейків. Дискримінатор спрямований на виявлення специфічних функцій події і збереження спільних для подій характеристик. Так відбувається перевірка новини на фейковість на базі контекстних даних [2].

Способи що ґрунтуються на знаннях ґрунтуються на факті, що новини частіше за все вміщують кілька складників: зображення, текст і метадані. Дані про подію не фіксується повною мірою окремою частиною. Окрема складова несе в собі лише частину загальної інформації. Кортежі мультимодальної інформації про новини можуть бути викривлені, якщо їх певна підмножина модифікується для подання або зміни мультимедійного набору. Зазвичай нюанси, які змінюють, щільно переплітаються з правдою. Тому способи основані на вмісті рідко виявляють маніпуляції такого роду.

Способи засновані на знаннях використовують набори готових даних в якості джерела знань, щоб перевірити **семантичну цілісність мультимедійних новин**.

Jaiswal вперше формально визначили проблему оцінки семантичної цілісності мультимедіа та об’єднали вивчення глибокого мультимодального представлення, щоб оцінити, чи відповідає підпис зображенню. Пакети даних еталонного набору були обрані для навчання моделі глибокого

мультимодального подання. Ця модель дозволила оцінити цілісність пакетів шляхом виведення балів відповідності надписів рисункам і застосування моделей виявлення вкидів для встановлення їх відповідності еталонному набору.

Проблемою створення методів оцінки семантичної цілісності мультимедіа є **відсутність даних для навчання та оцінки**. Для її вирішення було введено структуру Adversarial Image Repurposing Detection (AIRD). AIRD шукає зміни призначення зображень. Структуру можливо навчити за відсутності потрібних навчальних даних, що включають зманіпульовані метадані. AIRD моделює взаємодію між виконавцем, який використовує зображення з підробленими метаданими, і аналізатором, який перевіряє семантичну узгодженість між зображеннями та їх метаданими. AIRD складається з двох моделей: фальсифікатора та детектора, які навчені змагальним способом. Поки детектор збирає докази з набору посилань, фальсифікатор використовує їх, щоб створити переконливо оманливі фейкові метадані для даного пакета запиту.

На жаль, переглядаючи інформацію ми не завжди маємо можливість її перевіряти. Проте ми маємо право отримати правду, сепаруючи себе від неправди. Для цього необхідно знати правила, які дозволять зробити перші припущення про неправдивість інформації.

По-перше, варто вивчити варіанти пошуку фейків у друкованих засобах, адже вони не підлягають одразу перевірці за допомогою спеціалізованих програм чи інших автоматизованих інструментів аналізу. Переглядаючи джерело, ми спочатку акцентуємося на **заголовку**. Саме він формує перше враження, відповідно, зароджує або ні підозри новини у її фейковості. Більшість друкованих видань наповнені сенсаційними заголовками, які описують мало ймовірні факти. Проте чи співпадає сенсація у заголовку з вмістом публікації? Це перший прапор, що сигналізує про можливі підробки фактів.

По-друге, зосереджуємося на **ставленні автора до описаної події**. Коли автор стає на одну зі сторін конфлікту, при цьому активно засуджуючи іншу, -

це може бути маніпуляцією нашими почуттями.

Подібні маніпуляції часто спостерігаються під час проведення виборчих кампаній з метою швидкого отримання прихильності аудиторії через публікації від імені або за сприяння відомих і впливових людей, думкам яких довіряє суспільство.

По-третє, - це **джерела інформації**, наприклад, автори фотографій, імена людей, що займалися підготовкою матеріалів публікації, вказані дати тощо. Якщо джерела відсутні, - відразу повинна виникати підозра у недостовірності інформації, адже автори відразу обмежують нам шлях до її перевірки.

Ті самі правила розповсюджуються й на ту інформацію, яку ми отримуємо з телевізора або Інтернету, але додаються інші варіанти чекапу. *Наприклад*, **тональність репортажу**. Агресивний репортаж спрямований на виклик емоцій у глядача, а також на те, щоб висвітлити будь-яку корисну інформацію для того чи іншого медіа ресурсу у певному потрібному ракурсі. Сприймати інформацію потрібно спокійно й врівноважено, контролювати себе, й намагатися не піддаватися на психологічні прийоми досвідчених експертів, з ясним розумом набагато простіше аналізувати дані.

Висновки до розділу 1

В розділі описано суть поняття фейк. Виконано огляд історії появи цього поняття та небезпеки появи фейкових новин для суспільства. Розглянувши різні відомі класифікації фейків, ми вже можемо сформувати певні припущення, переглядаючи джерела даних на можливість вкладення у них фейкової частини. Проте автори з корисливих міркувань часом прикладають неймовірні зусилля, щоб якнайкраще приховати і затінити неправдиву або не зовсім правдиву інформацію, щоб у людей не з’явилося жодної підозри. Тому далі потрібно розбиратися зі способами діагностування фейків у потужному масиві інформації, яку ми щодня отримуємо з різноманітних засобів інформації.

Додатково у розділі було виокремлено головні характеристики, за котрими

типовий користувач інфосередовища буде спроможний сепарувати сфальсифіковані новини і правдиві. Це напрочуд корисно для підвищення довіри суспільства до джерел даних.

РОЗДІЛ 2. МЕТОДИ ВИЯВЛЕННЯ ФЕЙКОВИХ ПРОФІЛІВ КОРИСТУВАЧІВ

2.1 Поширеність фейкового контенту в соціальних мережах

Без перебільшення можна стверджувати, що однією з головних тем, що обговорюються професійними журналістами та дослідниками медіа є фейкові новини. У редакціях журналів, на потужних наукових семінарах, науково-практичних конференціях точаться гострі дискусії про те, як відрізнити фейк від правди та напівправди та про необхідність позначати новини, не підтверджені фактами. Це може бути дуже осудливо. Нік Ньюман у відомому рев'ю трендів 2017 року стверджує, що фейки підірвуть демократію і порядок у всьому світі. Генеральний директор Apple Тім Кук пропагував урядам усіх країн розпочати кампанію проти фейкових новин, що "вбивають людський розум". Разом з тим існують фахівці, які вважають фейки просто фейками, а саму цю проблему занадто перебільшеною. Так, наприклад, Том Узанузевскі, один із менеджерів інформаційного агентства Associated Press, заявив, що фейкові новини купує рекламодавець. По факту він спонсорує їх та заохочує розповсюдження. Найкращою відповіддю на фейкові новини є створення якісного матеріалу на ту ж тему з відповідними доказами та спростуваннями.

Щоб зрозуміти, чи загрожує брехня професії журналіста, потрібно визначити власне поняття фейку, що саме по собі є досить складною проблемою. Проблема полягає в тому, що поняття "**підробка"** стало широко використовуватися для позначення фотографій, оброблених у Photoshop, а іноді й відео, відредагованих у відеоредакторі, як підробки; сюди ж відносяться сторінки соцмереж, створені від імені інших людей (звісно, зазвичай відомих); анекдотичні історії про так звані видовища та джерела розваг.

Сьогодні **фейком** можна назвати будь-яку підробку. Але якщо говорити про інформацію та новини, то це цілеспрямоване використання сфабрикованих та підготовлених новин, основною метою яких є дискредитація будь-якої

установи, організації чи особи. Максимально близькими синонімами фейкових новин можна вважати **дезінформацію** або **брехню**. Творець фейкових новин має намір щось дискредитувати або когось збентежити.

Хоча інформаційні агентства повинні відмовлятися від фейкових новин, вони оперують психологією сприйняття фейкових новин як маніпуляції. Незважаючи на фальсифікацію документів і, власне, досвід, фейкові новини притаманні сучасній цивілізації з вірусними публікаціями в соціальних мережах. Цикл цифрових комунікацій став підходящим середовищем для популяризації фейкових новин. Містифікація як феномен перетворилася на найбільшу перевагу для Facebook, Reuters, Bloomberg та інших інформаційних середовищ.

Типовий користувач інформаційних мереж не володіє медіа грамотністю і не може відрізнити фейкові новини від справжніх. Згідно з даними відомої компанії Pew Research, кожне четверте повідомлення, яким американці діляться в Інтернеті, не є надійним. У сучасному світі дезінформація схожа на вірус, оскільки вона швидко привертає увагу читачів і поширюється.

Для тих, хто поширює фейкові новини, краще, щоб по можливості більше людей поглинули дезінформацію та поширили повідомлення на сайті новин. Для фейків важливо наповнювати свідомість користувачів фейковим змістом, який спотворює реальність і зароджує сумніви стосовно правдивості наявних фактів. Людина, яка наповнилася брехнею, втрачає імунітет до неї і робиться байдужою до різниці між добром і злом.

Монополії на правду не існує! Деяким користувачам мережі здається, що вони поширюють правдиві новини. Інші вважають, що, поширюючи неправду, вони кидають виклик консервативним поглядам і сприяють прогресу. Є джерела, що називаються сайтами сатири і займаються поширенням вигадок, позиціонуючи себе як таких, що борються з політичними режимами [6].

Але професійне співтовариство комунікаторів і самі соціальні мережі намагаються боротися з авторами фейкових новин і їх продукуванням. Facebook почав бачити фейкові новини у дописах користувачів. Зокрема, Seattle

Tribune News стверджують, що смартфон Дональда Трампа спричинив останні витоки інформації з Білого дому в пресу. Такі звіти супроводжуються надписом «суперечлива заява» у звіті Politifact і Snopes.com. Обидві спеціалізовані організації досліджують повідомлення після надходження запитів від користувачів. Наразі незрозуміло, чи буде більше звернень людей, чи лише дві перевірені організації будуть залучені до перевірки новинних повідомлень.

Зрозуміло, що найефективнішим заходом боротьби з брехнею є розвиток медіа грамотності серед мас. Користувачі соціальних медіа повинні відрізняти фейкові новини, які приносять користь комусь конкретно, і утримуватися від їх поширення. Тоді наше інформаційне суспільство не зможе заразитися цією інфекцією і стане несприйнятливим до подібних паразитів соціальних мереж [6].

Найбільш поширеною дезінформацією в соціальних мережах є повідомлення про COVID-19. Різноманітна інформація про агресивний вірус заповнила світові ЗМІ та інтернет-видання. Публікації часто згадуються у коментарях читачів Інтернету, популярних блогах тощо. Об’єктом розслідування є швидке поширення фейкових повідомлень про COVID- 19 у соціальних мережах, відтоді весь світ потрапив до карантину, саме Інтернет відіграв суттєву роль у розголосі інформації, включно з різноманітними чутками.

Згідно з даними схеми керування репутацією SCAN, розробленої в «Інтерфакс», у топ-5 «альтернативних» варіантів історії появи вірусу входять неправдиві повідомлення про те, що COVID-19 виготовили американці (1727 постів); COVID-19 розроблено китайськими вченими (1218 постів); COVID-19 - джерело слухань 5G (717 постів); COVID-19 прилетів із космосу (350 постів). Ця та подібні публікації націлені на підвищення інтересу аудиторії до інформації, прямо чи опосередковано пов’язаної з пандемією. Під час розслідування фейкових новин виявилося, що справжній намір автора фейку: «Я хочу привернути увагу аудиторії, тому я це говорю».

«Найкращі, найпереконливіші історії про коронавірус, незалежно від їх

достовірності» відповідають очікуванням заявників перед текстом («Я хочу отримати якомога більше інформації про COVID-19 і якомога швидше», «Я хочу отримувати інформацію з авторитетних джерел, медіа та Інтернет» і «Я довіряю інформації, що поширюється в ЗМІ та Інтернеті»). Інакше кажучи, наміри фейкових авторів і наміри споживачів отримати інформацію створюють фейкові новини і сприяють їх поширенню.

Зібрані та перевірені фейкові новини можна класифікувати відповідно до запитів на інформацію: хто є винуватцем або хто є джерелом інфекції? Яка статистика поширення коронавірусу? Хто зі знаменитостей інфікований? Як перевірити, чи здоровий ти? Яких дій шахраїв і лиходіїв чекати громадськості? Які заходи вживає влада проти інфекції?

У відповідь на ці та подібні запити здійснювалися перенаправлення на явно або відверто шахрайські сторінки. Іншими словами, фальсифікований матеріал — це відповідь на інформацію, яку запитує адресат, із посиланням на авторитетні джерела. ЗМІ та Інтернет свого часу вирували повідомленнями про те, що «перший китаєць заразився коронавірусом від бліх, які вільно продаються для споживання на ринках Уханя». Така новина цілком задовольняла потреби адресата, адже існує міф стосовно того, що китайці їдять будь-яку тварину, особливо якщо вони не піклуються про дотримання санітарних норм. Пізніше ця інформація почала широко поширюватися світом, а публічне звернення надало відповідь на хвилююче питання [6].

Однак очікується, що китайський адресат отримає інформацію про інше джерело інфекції, так як існуюча схема вживання продуктів у Китаї ніколи не викликала інфекцій такого типу. Як реакція на такі очікування незабаром з’явилася новина про те, що вірус міг бути привезений до Уханя американськими військовими. Стосовно цього 12 березня 2021 р. повідомив офіційний представник МЗС Китаю.

У своєму Twitter китайський дипломат Чжао Ліцзян включив відео виступу директора Центру з контролю та профілактики захворювань США (CDC) Роберта Редфілда перед Конгресом. На відео Р. Редфілд заявив, що у деяких

американців, які вважалися померлими від грипу, після смерті виявили COVID-19. Для тих, хто чекає підтвердження теорії змови, є фейкова новина про те, що COVID-19 штучно культивували в секретних бактеріологічних лабораторіях США: «Американські військові вірусологи експериментували з ранами і синтезували небезпечний вірус. Вони почали робити це, щоб виграти торгову війну з Китаєм».

Варіант теорії змови: «Пити з-під крана — не можна!» Фейкові новини показують, що водопровідну воду таємно отруюють, щоб сприяти зараженню. В соцмережах і ЗМІ поширювалася інформація про те, що з гелікоптерів будуть розпилювати дезінфекційні засоби. Крім того, різні версії цієї містифікації уточнюють деталі та подробиці: «Сьогодні з 11:00 вечора і до 5:00 ранку вертольоти будуть розпилювати дезінфекційний засіб, вікна та балкони мають бути закриті», «Не можна виходити після 10:00 ранку, як повідомили у військовому відомстві». У цій версії сказано: «Ніхто не повинен бути на вулиці об 11:00 сьогодні ввечері. Двері та вікна мають бути закриті, оскільки 5 вертольотів розпилюють дезінфікуючий засіб, щоб знищити COVID-19. Поширити цю інформацію для всіх контактів» [8].

Як ми знаємо, якась частка людей традиційно більше вірить народним засобам. Таким заявникам надсилають шахрайські повідомлення, в яких пропонують різноманітні традиційні методи профілактики інфекції або її лікування. Від початку у соціальних мережах почали пропонувати такі народні засоби: спирт, дитячу сечу, трав’яні збори, часник тощо. У березні в WhatsApp було поширене голосове повідомлення відомого лікаря Леоніда Рошаля, що вранці потрібно їсти часник. Інформація надійшла, коли Л. Рошаль розповідав про наслідки цих мір на лекції для студентів-медиків.

Як бачимо, зміст фейкових новин дуже різний і являє собою багато запитів споживачів щодо інформації про щось конкретне, *наприклад,* - коронавірус.

Організація фейкових розмов спрямована на тип адресата, який безсумнівно довіряє авторитетному джерелу в Інтернеті. Це жінки, які в першу чергу, турбуються про здоров'я близьких родичів; матері, які турбуються про

освіту своїх дітей, та інші категорії користувачів, які використовують Інтернет як джерело для отримання найновішої інформації.

Залежно від бажаної рецептивно-інтерпретаційної активності при отриманні повідомлень можна виділити два типи приймачів фальшивого контенту: раціонально-логічне та емоційно-емоційне.

Ще одна група, яка поширює дезінформацію про щеплення від COVID-19, — це противники вакцинації. Твердження кампанії базуються на не підтверджених наукою дослідженнях, не мають достатнього підґрунтя для підтвердження достовірності результатів, не вміщують епідеміологічних даних тощо. Вони використовуються для підтвердження власних висунутих гіпотез, поширення яких тісно пов’язане з використанням Інтернету, в основному через можливість швидкого поширення інформації та різних теорій (без епідеміологічних даних). У березні 2020 р. такі технологічні корпорації як Facebook, Google, LinkedIn, Microsoft, Reddit, Twitter і YouTube виступили зі спільною заявою стосовно протидії дезінформації про коронавірус [7].

У червні 2020 р. Facebook посилив рекламну політику, увівши табу на рекламу, яка обіцяє ліки від COVID-19. Через місяць генеральний директор Facebook Марк Цукерберг оголосив, що соціальна мережа надасть Всесвітній організації охорони здоров'я (ВООЗ) можливість вільно розміщувати інформацію, яка стосується розповсюдження COVID-19.

Facebook також розпочав оприлюднювати підтверджену інформацію про коронавірус COVID-19 у новинах. «Дата-центр» наповнений інформацією Всесвітньої організації охорони здоров'я і Центру контролю за захворюваннями США. Стосовно України, тут також повідомили МОЗ, так як раніше відомство уклало договір з Facebook про проведення відповідної кампанії.

Разом з тим Facebook скоригував спосіб взаємодії з юзерами, які зустрічаються з шахрайством. Компанія ініціювала розсилку повідомлень тим, хто лайкав, коментував або ділився неправдивою інформацією, яку видаляли за порушення умов обслуговування соціальної мережі. Повідомлення: «Ми видалили вашу улюблену публікацію, оскільки вона містить неправдиву та

потенційно небезпечну інформацію про COVID-19». Клікнувши на повідомленні, можна перемістится на спеціальну сторінку, на якій можна віднайти пост із поясненням причини його видалення з Facebook.

У березні було анонсовано, що всі публікації у Facebook та Instagram, які піднімають тему безпеки вакцини від COVID-19, будуть позначатися. У тексті зазначається, що всі вакцини обов'язково проходять тривале тестування на безпечність і дієвість до того, як отримати схвалення спеціалізованими регуляторними органами в усьому світі.

Facebook запустив новітню систему, з використанням якої користувачі можуть сказати, що підтримують вакцину від COVID-19.

Теми, які Facebook блокував поки тривала пандемія, включають: заперечення само ізоляції; ігнорування карантинних дій; недостатню ефективність вакцин проти цільових захворювань; хвороби безпечніші за вакцини; небезпечність, токсичність та аутичність вакцин.

Але не все ідеально в алгоритмах соціальних мереж. Зрештою, у цей період Facebook неодноразово звинувачували у фальсифікації достовірних записів і поширенні фейків.

У відповідності з дослідженням правозахисної групи Awaas, десять найпопулярніших сторінок Facebook, що поширюють фейкові новини та теорії змови, набирають у кілька разів більше переглядів, аніж десяток найліпших сайтів із доказами. Для прикладу, 2020 р. 82 веб сайти, що містять неправдиві відомості стосовно здоров'я, набрали 3,8 млрд звертань до сторінок. Найбільш відомими є 42 сайти з «гіперрозповсюдженням». Вони мають 28 млн передплатників, тільки за квітень 2020 р. було нараховано 800 млн звертань [7].

Алгоритми Facebook весь час змінюються, у тому числі для забезпечення контролю інформації. Неможливо все передбачити і протестувати наперед. Внаслідок цього у певний період було ненавмисно заблоковано інформацію, що рекламує вакцинацію проти COVID-19. Алгоритми соцмереж некоректно визначили цю рекламу як політичну. Інформація з Центрів контролю та профілактики захворювань (ЦКПЗ), Каліфорнійської медичної асоціації також

була заблокована помилково. Потім Facebook підтвердив, що некоректно розпізнавалися окремі оголошення.

У квітні 2020 р. Twitter змінив політику стосовно COVID-19. Було заплановано вилучення твітів із хибними обвинуваченнями відносно «підбурювання людей до дій та спричинення масової паніки, соціальних заворушень або загальних заворушень». Twitter запровадив інформаційне віконце на старті стрічки новин. Воно має надавати користувачам правдиву інформацію про вакцинацію від коронавірусу в їх країнах [6].

Натиснувши на лінку у вікні, читач переміститься на окрему сторінку в Twitter з даними стосовно вакцини, що відображається в Колекції твітів від відомих організацій ВООЗ та Центрів контролю та профілактики захворювань. У так званому твіттер-гіді є кілька розділів: наслідки щеплення, можливі побічні ефекти, поради для вагітних тощо.

Twitter також доповнив алгоритм функцією протидії поширенню дезінформації та фейкових новин, попереджаючи юзерів про підозрілий зміст, коли він їм подобається.

Раніше соцмережа видавала схожу нотифікацію при намірах переписати пост із суперечливою інформацією. Згідно зі статистикою компанії, це знизило розповсюдження хибних даних на 29%.

Twitter запустив пробний проєкт Birdwatch для боротьби з дезінформацією в соціальних мережах. Крім того, у квітні 2021 р. Twitter анонсував, що співпрацює з Associated Press (AP) і Reuters для противаги дезінформації через рекламу автентичного контенту.

У березні 2020 р. YouTube розпочав просувати офіційні повідомлення про COVID-19. На сайті з'явився спеціальний фрагмент з даними відносно коронавірусу COVID-19. Його мета – надавати офіційні новини та іншу інформацію про пандемію від відповідних органів. У травні 2020 р. YouTube видав політику щодо неправдивої медичної інформації про COVID-19. Зокрема, ці обмеження стосуються матеріалів, які містять викривлені дані стосовно COVID-19 з офіційних джерел (таких як Всесвітня організація охорони

здоров’я та місцеві органи охорони здоров’я). Тематика, охоплена новими правилами, включає ліки, профілактичні заходи, діагностування та способи поширення вірусу.

У разі розміщення такого вмісту Video Hosting спочатку повідомить вас електронною поштою про те, що його було видалено. Далі буде винесено друге попередження, а внаслідок 3-х попереджень вміст заблокується.

У жовтні 2020 р. YouTube почав блокувати контент, який пропагує теорії змови стосовно вакцини проти COVID-19. Було заявлено, що компанія покладатиметься на дані відносно вакцини від ВООЗ, Центрів з перевірки й попередження захворювань та експертів місцевих органів охорони здоров’я з усього світу [6].

У квітні відео-платформа YouTube анонсувала серію громадських послуг для вакцинації від COVID-19. У 16-секундному та 31-секундному відео [7] користувачам платформи рекомендується «повернутися до місця призначення» після ін’єкції. Проект запроваджується Платформою у партнерстві з Лондонським трестом вакцин здоров’я та тропічної медицини. У серпні 2020 р. Google ввів табу на використання його рекламної платформи з мето демо реклами рядом з контентом, який висвітлює теорії змови щодо COVID-19. Google також не дозволяє рекламу таких теорій.

У випадку оприлюднення сайтом матеріалів, що порушують ці правила, Google буде блокувати сайту використання його рекламної платформи.

У той час компанія Alphabet Google заборонила рекламу, що містила суперечливі заяви стосовно профілактики хвороб і неперевірені ліки, включаючи кампанії проти вакцинації або заохочення користувачів відмовлятися лікуватися.

У січні 2021 р. Google News запустила глобальний фонд протидії дезінформації про вакцину від COVID-19. Цій ініціативі було виділено близько трьох млн доларів США.

Фонд планує підтримувати ініціативи, націлені на перевірку фактів і зміну аудиторії для груп, що зазнали збитків від неправдивої інформації.

Повідомлення розглядатиме журі з 14 членів з наукового, медійного, медичного та некомерційного секторів, а також представників Всесвітньої організації охорони здоров'я.

Telegram досі залишається однією з соцмереж, яка не запровадила ніяких методів боротьби з дезінформацією, як-от блокування фейкової інформації про COVID-19. Фундатор Telegram стверджує, що організація не буде блокувати фейкові стосовно коронавірусу, так як цензура часто ускладнює противагу непорозумінням [6]. У січні 2020 р. Facebook, Twitter і Google анонсували об'єднання дій, спрямованих на попередження поширення дезінформації відносно вакцини від COVID-19. Компанії будуть співпрацювати з аудиторами та держустановами у всьому світі.

«Всеосяжні стандарти звітності даних» планується розробити ними та урядовими установами Великобританії й Канади.

27 березня 2020 р. Facebook анонсував запуск програми фактчекінгу в Україні. Друзями Facebook є VoxCheck і StopFake, незалежна міжнародна мережа фактчекінгу. Потужний сплеск цікавості до стрімко зростаючого потоку фейкових новин зумовлений не тільки їх бурхливим розмноженням у сучасних ЗМІ та Інтернеті, а й ступенем впливу такого вмісту на свідомість мас [6].

Адресоцентричний підхід до опису фейкових новин такий: «Що хоче знати керівник?», «Які є типи адрес фейкових новин?». Це включає дослідження для вирішення їхніх проблем. Аналіз у цій сфері дозволяє визначити основні теми фейкових новин. З точки зору змісту, контент фейкових новин різний і відображає різні потреби споживачів в інформації про коронавірус. Вміст новин, націлений на сприйняття емоційно рухливого підвиду читача, зазвичай, містить певну кількість статистичних і цифрових відомостей; в ньому присутні відомі факти; при цьому міститься оцінювальна лексика, представлені заклики до дій; такий контент традиційно містить окличні або питальні конструкції тощо.

2.2 Способи розпізнавання та аналізу

Нині відомо багато різних варіантів ідентифікації фейків, особливо фейкових фотографій, що використовуються надзвичайно часто у різноманітних соцмережах та на веб сайтах, іноді через дефіцит зображень для описання характерної новини, адже наявність фото, відео та постерів суттєво розширює спектр емоцій, які викликаються у користувачів під час перегляду інформації і коригують її сприйняття тим чи іншим чином.

Іншою стороною ситуації є те, що історичні фото (відео) часто модифікуються і пропонуються як нові для роздмухування описаних подій. Це частіше стається з фотографіями з місць глобальних катастроф, природних катаклізмів, військових дій тощо. Такі фото можна легко переробити і видати одну ситуацію за іншу, замінивши окремих учасників, місце тощо. Це значно спрощує процес підготовки матеріалу і здешевлює його. В якості прикладу можна навести фото війни з Сирії, які були адаптовані певними медіа і розміщені в мережі Інтернет як фото-звіти про бойові дії на території України у 2014 р.

Використання адаптованих історичних фото є поширеною проблемою, тому завдання виявлення сфальсифікованої інформації залишається актуальним. Нині існують різні програми, що виконують пошук за фото і суттєво полегшують життя користувачам мережі. Самі прості серед них - це плагіни, які встановлюються як додатки на пошуковий браузер. При обранні сумнівного фото (відео), можна досить швидко знайти дату публікації і джерело, у якому воно використовувалося раніше, і, базуючись на цьому, прийти до певних висновків.

Прикладом подібного плагіну є RevEye, що використовується для Google Chrome і функціонує за схемою, описаною вище. Іншим прикладом, який реально використати у різних браузерах, є плагін Who stole my pictures, що був розроблений для перевірки проєктів публікацій на те, чи використовувалися вони іншими людьми або веб ресурсами. В цілому він має той же принцип дії як і описаний плагін для Google Chrome. Додатково існують окремі додатки, що реалізують пошук за геолокацією фото і датою його створення. Це допомагає у

проведенні слідчих експериментів, пошуку свідків подій, огляду заданої локації з метою збору необхідної інформації.

Розглянувши кілька подібних додатків, можна стверджувати, що вони використовують досить різноманітні варіанти для перевірки істинності або недостовірності інформації. Тому далі будемо уточнювати, які аспекти загалом потребують уваги у процесі аналізу новин, і які саме складові є першочергово важливими для спішної ідентифікації фейків.

2.2.1 Зображення і графіка у розпізнаванні фейкових профілів

Візуальна частина кожної новини є одним з базових елементів, на котрих ми повинні акцентуватися в процесі аналізу даних. Вона може засвідчити багато як про саму інформацію, так і про її належність до фейків. Фахівцями у галузі ідентифікації фейків за візуальною складовою було класифіковано їх за певними ознаками: криміналістичні, семантичні, смислові і статистичні.

Криміналістичні: пошук фактів фальсифікування фото (відео), які автор залишив по собі, адже змінити щось без сліду практично неможливо. Це бувають інтервали, залишені під час суміщення різних відео, зміна якості зображення на різних фрагментах відео, кольорова гамма, виконання інтерполяції, характеристика шумів тощо. Процес встановлення подібних маніпуляцій основується на використанні бінарних патернів на зображенні, що змінюються відповідно до застосованої над ними маніпуляції. Застосування нейронних мереж ускладнило роботу слідчих по фейковим новинам, адже ШІ нині активно використовується для продукування нових медіа з високою часткою правдоподібності. Здебільшого слідча експертиза у таких випадках будується на перевірці сигналів, спектральних характеристиках і кореляції. Проте так як здебільшого фальсифікат основується на підробці відео контенту з залученням людей, то припущення відносно фейковості робилися і на основі візуальної складової, а саме - на неприродній поведінці людини в кадрі, деформації його рис обличчя тощо.

2.2.2 Семантика розпізнавання фейкових профілів

Напрочуд вагомою ознакою, яка здатна відокремити фейкову новину від звичайної, виступає **семантична складова**. Як було зазначено раніш, автори фейкових новин часто наводять їх у стані, що викличе найбільше емоцій у читача. Тому візуальна компонента неймовірно вагома, адже саме вона керує почуттями. Одним із потужних інструментів для дослідження семантики зображень є згорткова нейронна мережа VGG(візуальна геометрична група) від CNN. Вона використовується для різного роду зображень, але також може бути використана для дослідження фейків. Ця нейронна мережа складається з тринадцяти згорткових шарів і трьох основних, а також об’єднання шарів між собою. Тому розглянемо глибше її архітектуру:

- Вхід, на який модель очікує подачу фото визначеного розміру, а саме 244×244;
- Згорткові шари, які розділяють окремі ознаки фото і всі зміни, накладені на нього;
- Об’єднання шарів для спрощення аналізу за рахунок зменшення кількості ознак;
- З’єднані шари, що вирішують задачі класифікації.

Окрім цієї у напрацюваннях CNN є інші, схожі моделі, здебільшого націлені на пошук фейкових новин. Вони працюють на основі об’єднання інформації, що базується на аналізі пікселів та частот. Шар, що відповідає за аналіз піксельної складової й аналізує семантику. Окремо з метою пошуку певних емоційних провокацій, що часто присутні у фейках і є складовою семантичної частини, використовують так звані розгорнуті мережі.

2.2.3. Статистичні ознаки, їх використання у розпізнаванні фейкових профілів

Наступною характеристикою, важливою для аналізу фейкових новин є **статистичні риси**. Простими словами статистична риса відповідає за кількість та варіативність зображень, що відповідають тій чи іншій новині. На нашу думку, видається досить нетиповим, коли при перегляді стрічки новин на якусь екстра неординарну подію ми знаходимо лише одне фото. Сумнівно, що журналісти, які вели репортаж, виконали тільки одну або дві фотографії на її підтвердження, або взагалі на різних ресурсах ми бачимо одні й ті ж фото, однаковий ракурс тощо. Тому завжди доцільно здійснювати певну статистику, а саме - ми можемо ввести статистичні функції, що здійснюють аналіз розподілу правдивих новин і фейків. Деталізуємо базові властивості, на яких основується виконання статистичного аналізу:

- **Кількісні відношення**: прикладом може бути співставлення кількості фото, вжитих для опису, із загальною кількістю додатків до описаних подій.
- **Відношення популярності**: дослідження будується на відомих картинках, вибраних із стрічки, які можуть мати багато різних реакцій, наприклад, лайків, і порівнянні їх зі звичайними дописами.
- **Аналіз типів зображень**: переглядаючи сайти з текстами, ми бачимо фото різних типів; часом те чи інше джерело пробує дотримуватися єдиного формату, проте буває зустрічаються абсолютно різні, відтак знаходячи максимально популярний формат у потоці, далі обираємо його за основу і аналізуємо таким чином інші фото.

Тут описано лише деякі статистичні прийоми. Крім того існують і більш глибокі й **наукові**:

- **Оцінка чіткості**: метод розбіжності Куллбака-Лейблера між окремими сетами фото – колекцією (збірка фото з окремих подій) та подією (конкретне фото з аналізованої події), аналіз проводиться за формулою:

$$VCS = DKL(p(w|c)||p(w|k)), \quad (2.1)$$

де $p(w|c)$ та $p(w|k)$ оцінюють частоту появи візуального зображення певного слова w, у колекції або у самій події відповідно.

- **Оцінка когерентності**: функція схожості застосовується для порівняння

фото з колекції події, що перевіряється, на схожість (когерентність).

- **Оцінка через гістограму розподілу**: аналізуючи події цільової новини, складають матрицю подібності фото, на основі неї отримують певний вектор подібності фото, і, спираючись на ці дослідження, можна побудувати гістограму розподілу подібності фото на рівні дрібної деталізації.
- **Оцінка багатогранності візуальної складової**: оцінюється певна візуальна різниця між фото з колекції, за основу береться найбільш популярне зображення або таке, що відображує новину більш точно, найкраще по якості, і на її підставі виконується ранжування ряду інших фото з колекції. Для цього маємо таку залежність:

$$VDS = \sum^{N} 1 \sum^{i} (1 - sim\,(xi, xj\,)) \qquad (2.2)$$

- **Оцінка через кластеризацію**: використовують метод кластеризації з метою оцінювання колекції фото аналізованої новини, для реалізації цього використовують алгоритм ієрархічної агломеративної кластеризації.

2.2.4. Аналіз фейкових профілів з урахуванням контексту і метаданих

Обов'язковою опцією встановлення фейковості інформації є її контекст. Відмітимо, що звірка на співпадіння тексту з фото часто не дає очікуваного результату, адже творці фейків зазвичай стараються підлаштовувати підписи й інформацію у своїх опусах. Тому цей метод переважно стосується оцінки інших факторів, от як вміст веб сторінки, де розміщено новину, її графічна частина і метадані фотографій.

По-перше, деталізуємо **метадані**. Це певна інформація про графічні файли, дату, час і місце їх створення, розмір, тип файлу тощо. Дійсно, зібравши їх, ми зможемо знайти стосовно довільного графічного вмісту, ймовірно, всі необхідні нам дані для наступної оцінки. Проте такі дані часто губляться після публікації і, відповідно, вже не зможуть надати нам очікуваного результату. Однак коли вони все таки є, це зможе бути вагомим доведенням фейковості матеріалу.

По-друге, окрім метаданих не менш дієво користуються

контекстуальними характеристиками, отриманими ззовні способом зворотнього пошуку фотографій. Зворотній пошук фото виокремлюється від традиційного пошуку, тому що приймає фотографії як початкові дані і видає перелік релевантних веб сторінок, що включають відповідне фото, його назву, описову часитиу і час розміщення. Послідовність можна автоматизувати і використати для величезного сету фото через API пошукових систем, як от зворотній пошук фото Google.

Таким чином, окреслимо такі функції контексту:

- **Часові проміжки** - визначаються як часові інтервали між оприлюдненням новини та першою публікацією графічного контенту. Функція корисна у випадку чекапу оригінальності графічного контенту. Якщо виміряний час перевищує порогове значення, імовірність походження візуальної складової від нерелевантної події зростає.

- **Схожість між новинами** - подібність новин встановлюється як схожість між заголовком публікації і текстами сканованих веб сайтів. Текстова частина веб сайтів корисна для аналізу оригінальності фото. Функція дозволяє перевіряти відповідність між текстом презентації та візуальним контентом.

- **Надійність платформи** – обчислює ступінь довіри до вихідної платформи, де розміщено візуальний контент. На основі сету з даними, веб сайту, який надає правдиву інформацію про більше 2700 медіа, кожну веб сторінку, яку видає зворотний пошук фото, було оцінено за наступними параметрами: висока, низька або змішана фактичність. Частка веб сторінок кожної категорії, отриманих через зворотній пошук фото, була обчислена як характеристика надійності для відповідної платформи.

2.2.5 Вплив візуального контенту на результати аналізу фейкових профілів

Раніше нами було всесторонньо деталізовано такі види візуальних функцій, як криміналістичні, семантичні, статистичні і контекстні, для

виявлення мультимедійних фейкових новин. Ці функції відображають характеристики візуального вмісту та зазвичай поєднуються на практиці для охоплення більшої кількості ситуацій, які можуть бути сфальсифіковані. Не менш важливо розглянути типові підходи, що використовують візуальний вміст для перевірки фейковості новин. Їх зазвичай класифікують на **підходи засновані на контенті**, і **підходи засновані на знаннях**.

Підходи засновані на контенті зосереджені на комбінації сигналів із контенту різних модальностей для встановлення фейковості новин за умов відсутності довідкових сетів даних.

Підходи засновані на знаннях зосереджені на використанні зовнішніх джерел для перевірки фактів на вхідних даних. Вони вимагають наявності досить великого набору еталонних даних, і оцінюють справжність публікації, порівнюючи її з відомими текстами, присутніми в еталонному сеті даних.

В цілому довільна новина складається з тексту і візуального супроводу. Такі складники надають певні підказки, використовуючи які на підтвердити фейковість. Новітні дослідження цієї проблеми зосереджувалися на комбінуванні інформації з різних модальностей. Переважно вони опираються на рекурентну нейронну мережу (RNN) і заздалегідь навчену CNN для встановлення текстових і візуальних семантичних характеристик і наступного дослідження.

Для найбільш сучасних підходів характерне комбінування наявної інформації з усіх зрізів для встановлення фейковості, тобто створення мультимодальної інформації. Мультимодальний контент включено через глибокі нейронні мережі для вирішення проблеми автоматизації перевірки новин. Було запропоновано інноваційний RNN із механізмом уваги для підвищення ефективності оцінки поєднання текстового, візуального і соціального контексту. Для кожної окремої новини текст і соціальний контекст спочатку поєднують з LSTM. Таке подання потім доповнюють візуальними характеристиками, які отримують із заздалегідь навченої глибокої CNN.

Вихідні дані LSTM на окремому часовому інтервалі використовують в

якості значення на рівні нейронів для співставлення візуальних ознак у процесі поєднання.

Окремо пропонувалася наскрізна нейронка для подій з мето пошуку нових фейкових новин на базі мультимодальних функцій, притаманних аналізованій події. Вона вміщує три базові складники: **мультимодальний екстрактор ознак, детектор фейків і дискримінатор подій**. Мультимодальний екстрактор ознак потрібен для виокремлення текстів і фото з публікацій. Екстрактор тісно зв’язаний з детектором фейків, адже потрібно навчитися дискримінаційному представленню з метою подальшого аналізу фейків. Роль дискримінатора полягає у витягуванні характерних події частин і зберіганні унікальних особливостей подій. Так виконують перевірку фейковості новини базуючись на **контексті даних**.

Підходи засновані на знаннях побудовані на факті, що мультимедійні новини традиційно складають з окремих частин, таких як фото (відео) з супровідним текстом і додатковими описовими даними, де інформація стосовно події не фіксується кожною частиною окремо, інакше кажучи кожна частина може надати нам лише окремі пазли про подію. Набори різнотипних даних або кортежі мультимодальної інформації по новинах, схильні до маніпуляцій, коли їх частина може бути замінена для подачі або заміни мультимедійного пакета. Деталі, які коригують, зазвичай є достатньо правдивими, тому підходи, побудовані на контенті, не можуть виявляти подібні маніпуляції.

У випадку виявлення такої проблеми підходи, засновані на знаннях, відразу використовують зовнішні джерела (набір довідкових даних необроблених пакетів) як джерело світових знань, щоб виконати чекап семантичної цілісності мультимедійних новин.

Jaiswal вперше описали питання перевірки семантичної цілісності мультимедійного вмісту і поєднали вивчення глибокого мультимодального подання без включення інших методів виявлення для оціни відповідності тексту фотографії. Набори даних еталонного набору було використано для навчання моделі глибокого мультимодального подання, яка у подальшому була

застосована для виведення оцінки цілісності пакетів запитів способом обчислення балів узгодженості текстів із фото.

Типовою проблемою у розробці методів оцінки семантичної цілісності мультимедіа традиційно являєтьсяє відсутність даних. Тому було запроваджено таку структуру як Adversarial Image Repurposing Detection для встановлення зміни призначення фото. Її можна навчити за відсутності навчальних даних, що містять змінені метадані. AIRD моделює реальну взаємодію між виконавцем, який використовує фото з підробленими метаданими, і аналізатором, котрий перевіряє семантичну відповідність фото і метаданих супроводу. Зокрема, AIRD вміщує такі моделі: **фальсифікатор** і **детектор.** Моделі навчені способом змагання. Детектор формує факти з набору лінків, у паралелі з цим фальсифікатор використовує їх для продукування правдоподібних фейкових метадаих.

2.2.6 На що потрібно звертати увагу під час ознайомлення з новою інформацією

Отримуючи публікацію ми не маємо можливості глибоко перевіряти її. Конституційно ми маємо право на правдиві дані. Інстинктивно ми огороджуємо себе від неправди вмикаючи механізм самозбереження. Відповідно, для автоматизації процесу аналізу даних ми повинні чітко визначити правила перевірки правдивості отриманих нами відомостей.

По-перше, потрібно вивчити способи виявлення фейків у засобах медіа, адже вони не одразу піддаються перевірці програмами чи іншим автоматизованим способів аналізу. Переглядаючи видання, ми спочатку звертаємо увагу на **заголовок**. Саме він формує підсвідому оцінку, а також є першочерговим способом виявлення або підозри на фейковість новини. Багато журналів та газет наповнені сенсаційними заголовками, що описують неймовірний факт або на чомусь наголошують та чи співпадає інформація у заголовку з тим, що описано у статті? Це перший маркер, який вказує на те, що

певна стаття може містити підроблені факти, бо що заважає авторам гучного заголовку написати ще більш гучні факти у статті.

Друге, на що маємо звернути увагу, - це **ставлення автора до події**, яка описана. Якщо автор погоджується з однією стороною конфлікту в описаній події і засуджує іншу, це може бути маніпуляцією у певну сторону. Переглядаючи матеріал ми наперед маємо певну довіру або недовіру до видання і автора. Відтак якщо підтримуємо його точку зору, то автоматично піддаємося свідомій маніпуляції. Такі прийоми часто помітні протягом виборчих кампаній. Просто отримати прихильність аудиторії через позицію відомих і впливових людей, які виступають авторами певного контенту в засобах медіа.

Третій важливий момент, на котрий ми зобов'язані зважати, - **джерела інформації**, *наприклад,* автори фото, імена дослідників події, дати публікацій тощо. Коли джерела відсутні, повинна виникати підозра у фальсифікації даних, адже її неможливо перевірити. Ті ж правила поширюються на інформацію, що надходить з глобальної мережі. Важливі також такі параметри як тональність репортажу. Не збалансований, агресивний контент традиційно виступає маркером для перевірки контенту. Аналізувати дані потрібно спокійно й врівноважено, щоб не піддаватися у процесі тонким психологічним прийомам досвідчених фахівців. З чистим розумом значно простіше знайти правду у контенті, що пропонується.

2.3 Використання machine learning для виявлення фейкових профілів користувачів

Машинне навчання суттєво полегшує життя користувачів, у тому числі за рахунок автоматизації процесу аналізу новин. Тому доцільно розглянути кілька кращих алгоритмів машинного навчання, придатних для пошуку фейкових новин.

1. Метод GNN

Відомі праці такі як Graph Neural Networks with Continuous Learning for Fake News Detection from Social Media [30] і User Preference-aware Fake News Detection [43], побудовані на комбінуванні подання речень на базі уваги і навчання на них графової нейронної мережі. Система аналізує текст і метадані користувача (*наприклад*, твіти або історію публікацій) не тільки для пошуку неправдивих новин, а також для аналізу власне користувачів (їх акаунтів), які потенційно генерують неправдиві новини, виходячи з їх типової поведінки.

2. Метод BiLSTM + увага

Метод зайняв друге місце в рейтингу FNC-1 (Fake News Challenge Stage -1) із середнім значенням точності 84,60. У статті "Поєднання ознак схожості та глибокого навчання репрезентації для виявлення позиції в контексті перевірки фейкових новин" дво-направлені рекурентні нейронні мережі (RNN) використовуються паралельно з максимальним об'єднанням часового/послідовного виміру і увагою подання заголовка, перших двох речень новини й усієї статті. Ці подання об'єднуються і передаються на кінцевий рівень, котрий будує прогноз для позиції новин поточного ЗМІ.

3. Метод CNN+DNN

Метод векторизує текст, закладаючи в основу термін частота-обернена частота документа (TF-IDF). Він знаходить схожість заголовків і основного тексту. Отримана матриця подається на вхід поверхневої штучної нейронної мережі (ШНМ) для вилучення ознак. Після цього подається на глибоку нейронну мережу з прямим поширенням, що завершується шаром softmax. Метод гарно спрацьовує для набору даних FNC-1, так як різниця між правдою і неправдою тут очевидна.

Однак термін "частота, обернена до частоти документа" (TF-IDF) і поєднання CNN і DNN допомагають розпізнати фейк.

4. Метод CNN+Boosted Trees

Комбінація CNN+Boosted Trees - це тип машинного навчання, котрий спрямований на виявлення фейкових новин. Метод суміщує можливості згорткових нейронних мереж (CNN) і дерев, що виокремлюють типові ознаки з

початкового тексту і класифікують його як фейк чи правду. У схемі потрібно спершу виконати попередню обробку вхідного тексту, його перетворюють у матрицю вкладених слів. Після цього до матриці застосовують CNN, щоб виокремити важливі ознаки. В результаті отримуємо вектор ознак, який подаємо у модель Boosted Trees. По суті це дерево рішень, яке містить кілька слабких класифікаторів для подальшого формування сильного класифікатора. Таким чином, базуючись на виокремлених ознаках Boosted Trees навчається розділяти новини на фейкові й правдиві.

Метод CNN+Boosted Trees має певніи переваг у виявленні фейків. Він може обробляти тексти довільної довжини, фіксувати їх контекст і зміст. Окрім того, він менш схильний до перенавчання, порівняно з іншими підходами машинного навчання. Поєднання CNNs і Boosted Trees дозволяє підвищити надійність і точність у пошуку фейків.

5. MLP

Метод MLP (Multilayer Perceptron) базується на нейронній мережі, яку можна використати для виявлення фейків за допомогою машинного навчання. Він складається з декількох шарів взаємопов'язаних вузлів і може навчатися на розмічених даних для подальшої класифікації новин як фейкових чи правдивих. Для виявлення неправдивих новини спочатку виконують попередню обробку вхідних текстових даних з метою їх перетворення у числове подання. Далі це подання використовують в якості вхідних даних для MLP, котрий навчається класифікувати новини як фейкові або правдиві на основі ознак, витягнутих із вхідних текстових даних. Процес навчання передбачає коригування ваг вузлів мережі для мінімізації похибки.

Метод MLP має переваги в керуванні складними, нелінійними зв'язками між входом і виходом. Коли навчальний набір даних досить великий і різноманітний, він також досить точно вирішує завдання класифікації. Такі популярні інструменти машинного навчання TensorFlow і Kerasна також можна використати для ефективної реалізації алгоритму MLP.

Висновки до розділу 2

Таким чином, у цьому розділі досить детально розглянуто різні варіанти розпізнавання і аналізу фейкових новин, їх специфіку, характерні переваги і недоліки. Для наступного дослідження буде використано метод, який основано на семантиці новини у поєднанні з машинним навчанням, тому що саме він є найбільш зручним для опрацювання якраз текстової інформації, що й буде підлягати аналізу в рамках цієї роботи.

Також у цій частині було виокремлено типові ознаки, за котрими типовий користувач інформаційного середовища може сепарувати новини на сфальсифіковані і правдиві, що є напрочуд корисним у контексті підвищення довіри людей до актуальних інформаційних джерел.

РОЗДІЛ 3. РОЗРОБКА МОДЕЛІ РОЗПІЗНАВАННЯ ФЕЙКОВИХ ПРОФІЛІВ

3.1. Структура облікових записів

Обліковий запис - збережена в комп'ютерній системі сукупність даних про користувача, необхідна для його розпізнавання (автентифікації) та надання доступу до його особистих даних і налаштувань. Як синоніми також використовується розмовне «аккаунт» (англ. *account* - «обліковий запис, особистий рахунок»).

Для використання облікового запису (іншими словами, для входу в систему під ім'ям певного користувача) зазвичай потрібне введення імені або логіну (англ. *login*) і пароля (англ. *password*). Також може вимагатися інша додаткова інформація.

Користувачі інтернету можуть сприймати обліковий запис як особисту сторінку, профіль, особистий кабінет, місце зберігання особистих та інших відомостей на певному інтернет-ресурсі (соціальній мережі).

Основними складовими облікових записів у соціальних мережах є ім'я користувача, його фотографія (аватар), коротка інформація про нього (*наприклад*, дата народження, країна, місто, інформація про освіту та роботу, інтереси тощо), коло спілкування користувача (друзі, друзі друзів, підписники тощо), альбом для фотографій. Оскільки головною метою соціальних мереж є спілкування з іншими користувачами, кожна соціальна мережа мас свою систему обміну повідомленнями, а також іншими файлами (зображення, відеозаписи, текстові документи тощо).

Часто зустрічається можливість об'єднування користувачів у спільноти за інтересами (групи) або за іншими ознаками, а також можливість розміщення публічних повідомлень та їх коментування.

Деякі соціальні мережі використовують фон облікового запису для унікального оформлення сторінки користувача, а також можливість слухати

музику та переглядати відеозаписи. Інколи користувачам надається можливість записувати свою контактну інформацію (номер мобільного телефону, адресу електронної пошти, логін у Skype тощо для організації обміну даними між окремими користувачами).

Так, на рисунку 3.1 зображено зовнішній вигляд облікового запису Марка Цукерберга у соціальній мережі «Facebook» [51], де можна побачити його ім'я, фотографію, інформацію про нього, список підписників та публічні повідомлення з прикріпленими фотографіями (пости).

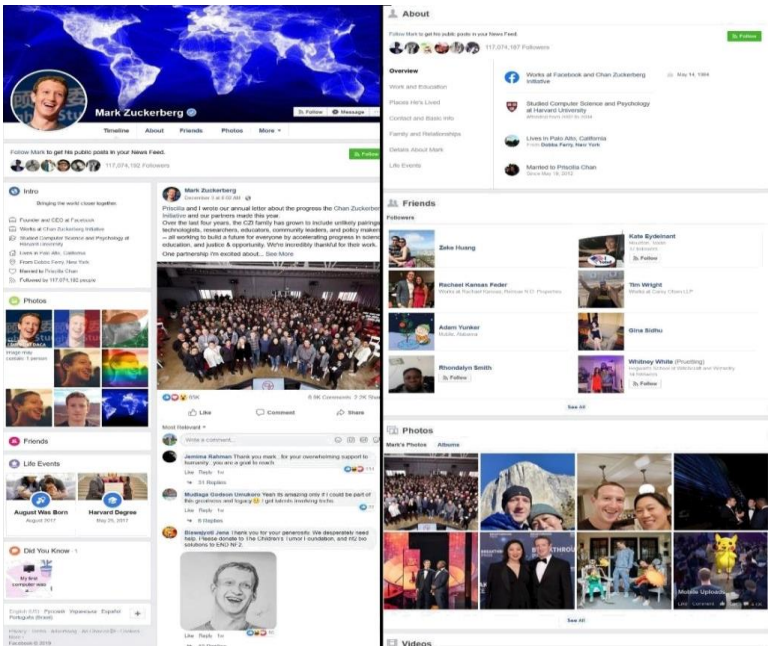

Рис. 3.1. Структура облікового запису на прикладі сторінки Марка Цукерберга у соціальній мережі «Facebook»

Обліковий запис зазвичай містить відомості, необхідні для впізнання користувача при підключенні до системи, відомості для авторизації і обліку. Це ідентифікатор користувача і пароль. Пароль або його аналог типово зберігається в зашифрованому або захешованому вигляді для забезпечення безпеки користувача. Проте, вже не є таємницею, що «Facebook» зберігає на своїх серверах величезну кількість інформації про облікові записи, яку не видно звичайному користувачу. До такої інформації належать [26]:

– дані про геолокації користувача;
– дані про коментарі й пости користувача;

- дані про відмітки користувача на фото тим чи іншим користувачем;
- потужність сигналу мережі;
- використану операційну систему;
- використаний браузер;
- оператора зв'язку / інтернет-провайдера;
- cookies третьої сторони;
- приховані користувачем публікації;
- вилучена з облікового запису інформація;
- відмітки «подобається»;
- багато іншої інформації.

Цю інформацію «Facebook» використовує для аналізу поведінки та побудови загальної статистики про користувачів, а також для формування рекламних повідомлень відповідно до вподобань користувача. Наприклад, для кожного окремого поста зберігається декілька мегабайт (від 3 до 8 мегабайт) інформації, у якій міститься інформація: про автора поста; кількість лайків та тих, хто їх залишив; про користувачів, що переглянули пост; текстова частина поста; інформація про зображення поста (якщо таке існує); дата створення поста; кількість користувачів, що поділилися постом; тощо.

Таким чином, загальну структуру облікового запису у соціальних мережах можна зобразити у вигляді схеми (рис. 3.2), яка відображає як видимий вміст облікового запису, який бачить користувач, так і невидиму звичайному користувачу інформацію про нього та його дії, що зберігається на серверах соціальної мережі.

Рис. 3.2. Загальна структура даних облікового запису у соціальних мережах

Необхідно пам'ятати, що не вся інформація про обліковий запис може бути доступною для перегляду, оскільки сучасні соціальні мережі мають можливість приховувати інформацію від сторонніх користувачів, при цьому інформація буде доступною лише друзям або обмеженому колу користувачів. Також, далеко не вся інформація про користувача є правдивою. Марк Цукерберг, засновник «Facebook», що є найбільшою соціальною мережею у світі, заявив, що близько половини облікових записів (а це більше 1 млрд облікових записів) є фейковими. Тому наразі виявлення фейкових облікових записів у соціальних мережах є актуальним завданням.

3.2 Метрики облікових записів у соціальній мережі

Дослідження показали, що можна виділити такі основні категорії ознак фейкових облікових записів: лайки, персональні дані, статуси та посилання, друзі, фото, дата народження.

Лайки (*LIKES*) за ознаками можна поділити на їх кількість *(QUANTITY)* та хто їх залишив на сторінці користувача *(FROM)*. У свою чергу залишити лайки можуть як друзі *(Friends),* так і незнайомці *(Strangers).* Для визначення фейковості профілю також має значення кількість лайків *(NumberOfLikes).* Якщо у користувача під певним постом кількість лайків більша за кількість його друзів *(NumberOfFriends),* це може свідчити про те, що користувач отримав ці лайки незвичним шляхом. Відсутність лайків *(NumberOfLikes =0)* на сторінці вказує на «ізоляцію» користувача, що також може свідчити про його фейковість. Структурна модель метрик у категорії «Лайки» показана на рисунку 3.3.

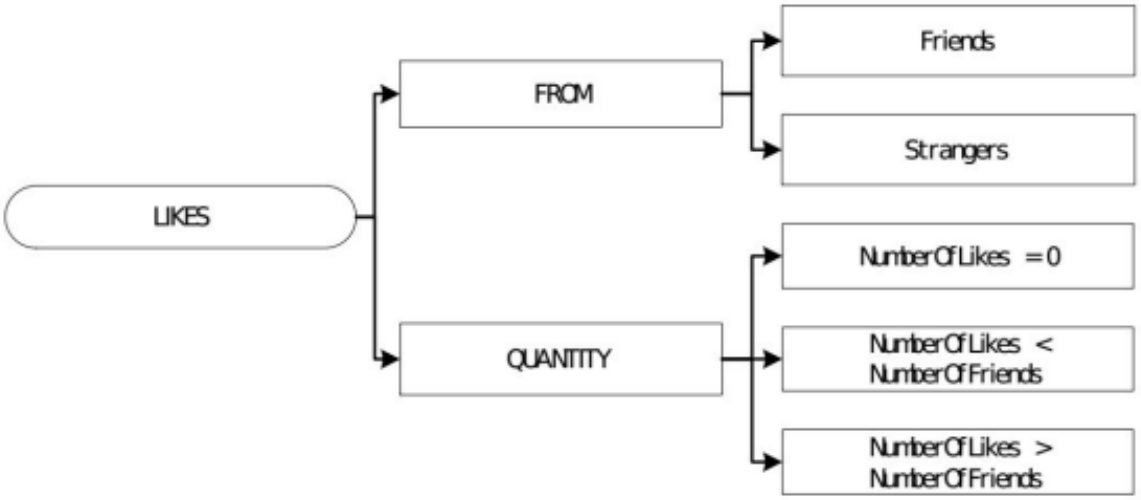

Рис. 3.3. Структурна модель ознак фейковості у категорії «Лайки»

Параметри моделі ознак фейковості у категорії «Лайки» можна записати у вигляді кортежів:

LIKES = {FROM; QUANTITY}

FROM = {friends; strangers}

QUANTITY = (NumberOfLikes = 0; NumberOfLikes<NumberOfFriends; NumberOfLikes>NumberOfFriends}

Персоналъна інформація на сторінці користувача PERSONAL INFORMA TION ABOUT USER) [23] може сказати чимало про фейковість або справжність профілю. Для подальшого аналізу персональну інформацію можна поділити на

дату народження *(DATE OF BIRTH)* ім'я користувача *(USER'S NAME),* кількість інформації *(QUANTITY OF INFORMATION),* суперечливу інформацію *(CONTRADICTORY INFORMATION)* та приватну інформацію *(PRIVATE INFORMATION).*

Дата народження має ознаки, які можуть вказувати на фейковість сторінки. Часто користувачі фейкових облікових записів не приділяють уваги детальному заповненню сторінки та залишають дату народження за замовчуванням (зазвичай 1 січня). Також можлива ситуація, коли вік користувача (*Age*) підлягає сумніву або не співпадає з іншими датами на сторінці *(DatesOnProfile).*

Ім'я користувача дослідити важко, оскільки існує чимало людей з таким самим ім'ям та прізвищем. Проте варто перевірити ім'я на предмет його співпадіння з ім'ям видатної людини *(IsCelebrityName).* Також слід звернути увагу на те, чи належить ім'я користувача *(UserNameCountry)* до типових імен країни цього користувача *(UserCountry).*

Відсутність персональної інформації у профілі *(NoInfoAboutUser)* або неповністю заповнені поля для персональної інформації *(PartlyFilledInfo)* свідчить про те, що користувач не хоче, щоб його могли ідентифікувати інші користувачі, а отже це теж є ознакою фейковості. Проте, існують також облікові записи, де персональна інформація заповнена повністю *(Fu11yFilledInfo),* хоча і недостовірними даними.

Суперечливість інформації на сторінці є одним з найбільш достовірних показників фейковості, проте це потребує складного аналізу. *Наприклад*, інформація в постах користувача *(Postslnfo)* не відповідає інформації, зазначеній у профілі *(InfoAboutProfile),* або користувач знаходиться у групах *(InfoAboutGroups),* які не відповідають його зазначеним інтересам *(InfoAboutIntersts).*

До персональної інформації також відносять номер телефону *(PhoneNumber)* та посилання на облікові записи у інших соціальних мережах *(SocialNetworksLinks).* Користувачі рідко виставляють таку інформацію у відкритий доступ, проте вона може свідчити про те, що користувач є активним

у соціальних мережах і не має намірів приховувати інформацію про себе. Також цим засобом можуть користуватися спеціально створені рекламні профілі.

Структурна модель метрик у категорії «Персональна інформація про користувача показана на рисунку 3.4.

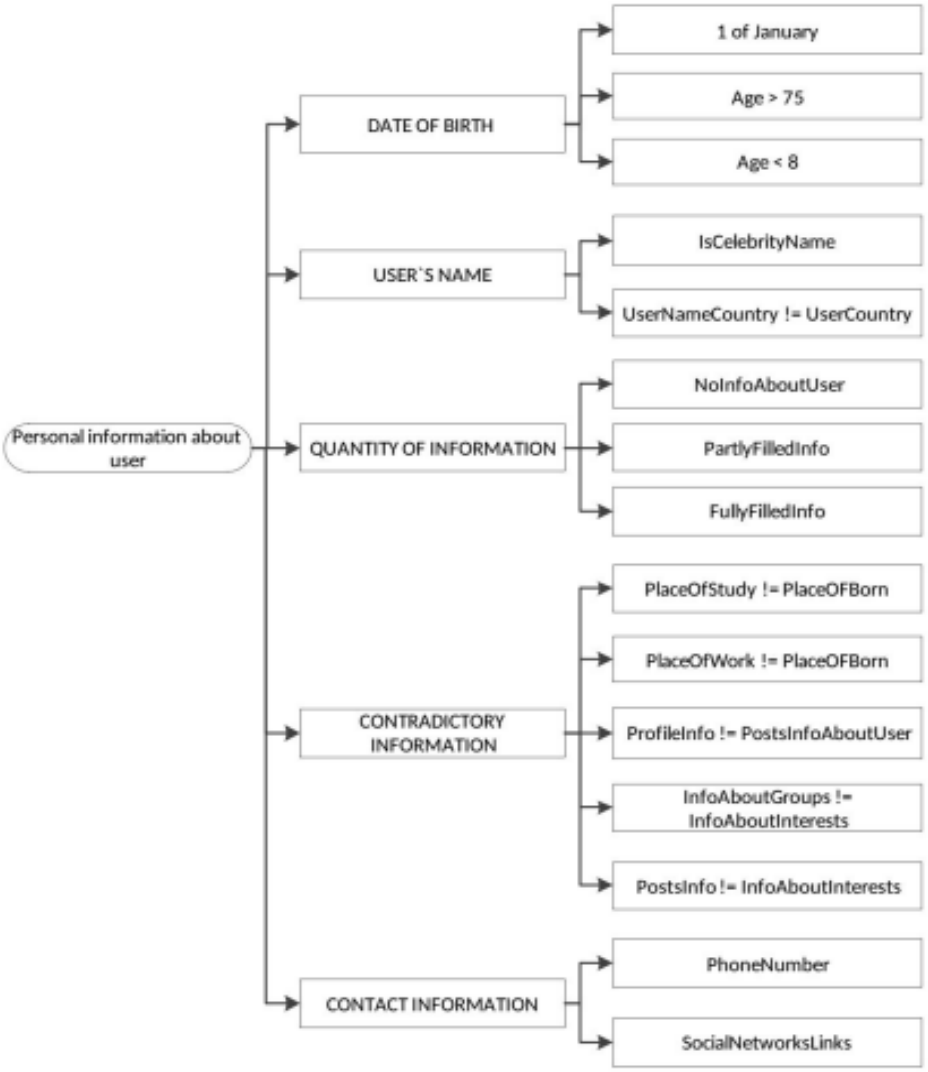

Рис. 3.4. Структурна модель ознак фейковості у категорії «Персональна інформація про користувача»

Параметри моделі ознак фейковості у категорії «Персональна інформація про користувача» можна записати у вигляді кортежів:

PERSONAL INFORMATION ABOUT USER = (DATE OF BIRTH; USER NAME; QUANTITY OF INFORMATION; CONTRADICTORY INFORMATION; CONTACT INFORMATION}

DATE OF BIRTH = {1 of January; Age}

USER NAME = {IsCelebrityName; UserNameCountry/UserCountry}

NUMBER OF INFORMATION = {InfoAboutUser; PartlyFilledInfo; FullyFilledInfo}

CONTRADICTORY INFORMATION = {PlaceOfStudy/PlaceOfBorn; PlaceOfWork/PlaceOfBorn; ProfileInfo/PostsInfoAboutUser; InfoAboutGroups/InfoAboutInterests; Postslnfo/InfoAboutInterests}

CONTACT INFORMATION = {isPhoneExists; SocialNetworksLinks}

Статуси та пости (STATUSES AND POSTS ON PAGE) на сторінці користувача аналізуються як одне ціле, оскільки вони відрізняються лише розміщенням у профілі. Їх можна аналізувати за такими ознаками: за частотою редагування/додавання *(UPDATES)* та за коментарями *(COMMENTS).* Статуси та пости іноді використовуються у якості реклами *(Advertising)* [31].

Частота редагування/додавання постів та статусів (*UpdateFrequency)* вказує на активність користувача. Якщо пости/статуси додаються рідко або дуже часто – це є однією з ознак фейковості. Якщо користувач давно додав пост/статус і протягом тривалого часу не оновлює, існує імовірність того, що цей обліковий запис є фейковим.

Кількість коментарів *(QUANTITY)* також вказує на активність самого профілю. Їх відсутність або надмірна кількість найчастіше буває саме у фейків. Коментарі можуть залишити *(FROM)* як друзі користувача *(Friends),* так і незнайомці *(Strangers).*

Структурна модель метрик у категорії «Статуси та пости» показана на рисунку 3.5.

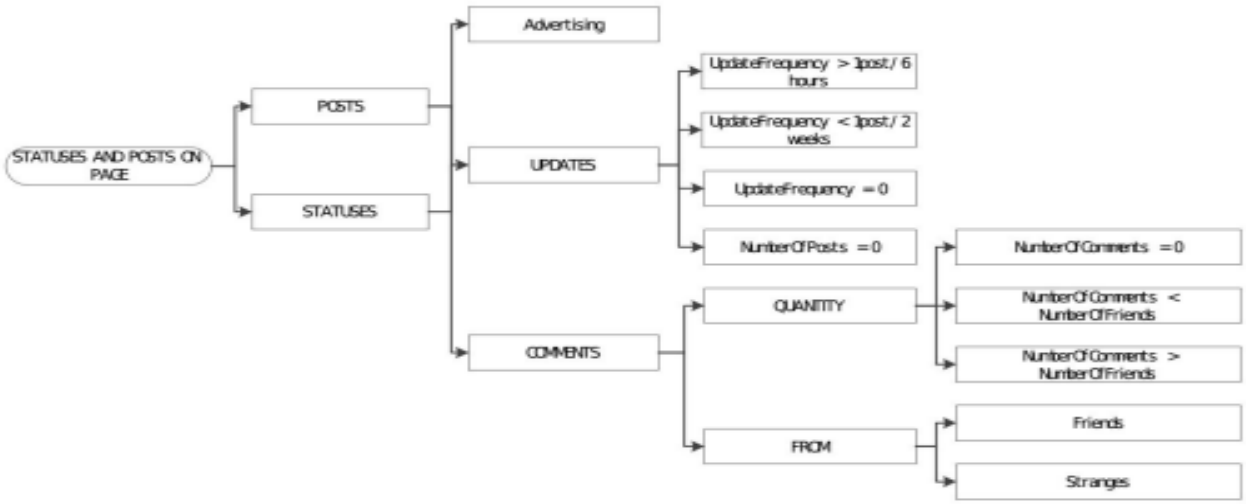

Рис. 3.5. Структурна модель ознак фейковості
у категорії «Статуси та пости»

Параметри моделі ознак фейковості у категорії «Статуси та пости» можна записати у вигляді кортежів:

STATUS AND POSTS ON PAGE = {POSTS; STATUSES}

POSTS = {Advertising; UPDATES; COMMENTS}

STATUSES = {Advertising; UPDATES; COMMENTS}

UPDATES = {UpdateFrequency; NumberOfPosts}

COMMENTS = {QUANTITY; FROM}

UANTITY= (NumberOfComments; NumberOfComments / NumberOfFriends}

FROM = {; Strangers}

Друзі користувача *(FRIENDS)* грають доволі значну роль у визначенні фейку, оскільки вони вказують як на активність профілю в соціальній мережі, так і на коло інтересів користувача [28].

Залежність фейковості від кількості друзів *(QUANTITY)* користувача проаналізувати важко, тому що для того щоб зробити висновок про фейковість, необхідно аналізувати самих друзів. Наприклад, якщо у списку друзів користувача є фейки *(IsFriendFake)*, є імовірність, що і сам користувач – фейк. Також важливо досліджувати зв’язки між друзями та друзями друзів *(LinksBetweenFriends)*. Якщо користувач не має друзів *(NumberOfFriends=0)*,

існує велика імовірність, що його профіль використовується не для спілкування, а для інших цілей. Велика кількість друзів *(Max(NumberOfFriends))* за короткий проміжок часу *(Min(timeline))* після створення профілю також викликає підозри, тому, скоріше за все, такий профіль є фейковим. Структурна модель метрик у категорії «Друзі» показана на рисунку 3.6.

Параметри моделі ознак фейковості у категорії «Друзі» можна записати у вигляді кортежів:

FRIENDS = {QUANTITY; INFORMATION ABOUT FRIENDS}

QUANTITY = {NumberOfLikes; NumberOfFriends; NumberOfLikes/ NumberOfFriends; Max(NumberOfFriends); Min(timeline)}

INFO ABOUT FRIENDS = {IsFriendFake; CommunityOfFriends; LinksBetweenFriends}

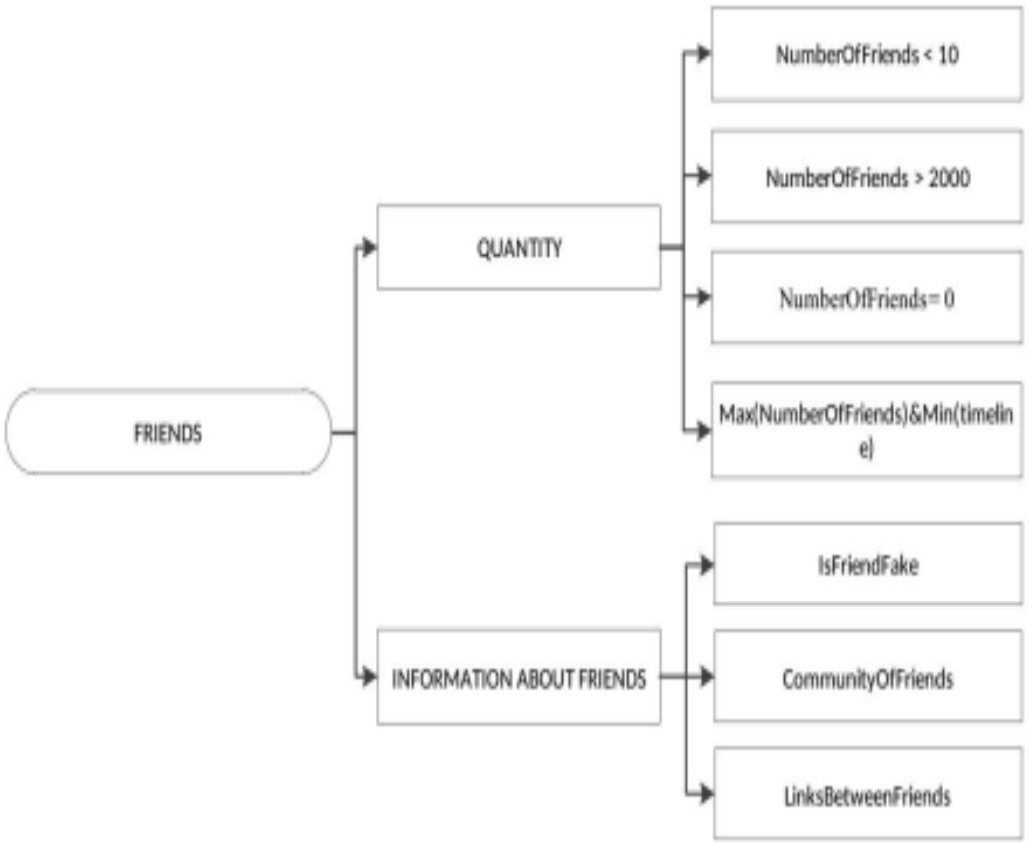

Рис. 3.6. Структурна модель ознак фейковості у категорії «Друзі»

Аналіз *фотографій користувача (PHOTO)* відіграє найважливішу і,

водночас, найважчу частину дослідження фейковості облікового запису. По-перше, необхідно проаналізувати присутність фотографій на аватарі (EXISTANCE). Відсутність фотографій як на аватарі *(AVATAR / COVER),* так і в альбомах *(PROFILE)* вже свідчать про те, що даний обліковий запис є фейковим. По-друге, за наявності фотографій на сторінці їх потрібно аналізувати на предмет співпадіння з іншими фотографіями в Інтернеті або з фотографіями інших профілів, оскільки користувач міг завантажити замість своїх фотографії знаменитостей *(Celebrities)* або інших об'єктів *(OtherPictures).* Кількість фотографій *(HumberOfPhotos)* також є важливим показником, оскільки надмірна або замала кількість фотографій вказує на неправдивість фотографій або неактивність користувача, відповідно.

Структурна модель метрик у категорії «Фото» показана на рисунку 3.7.

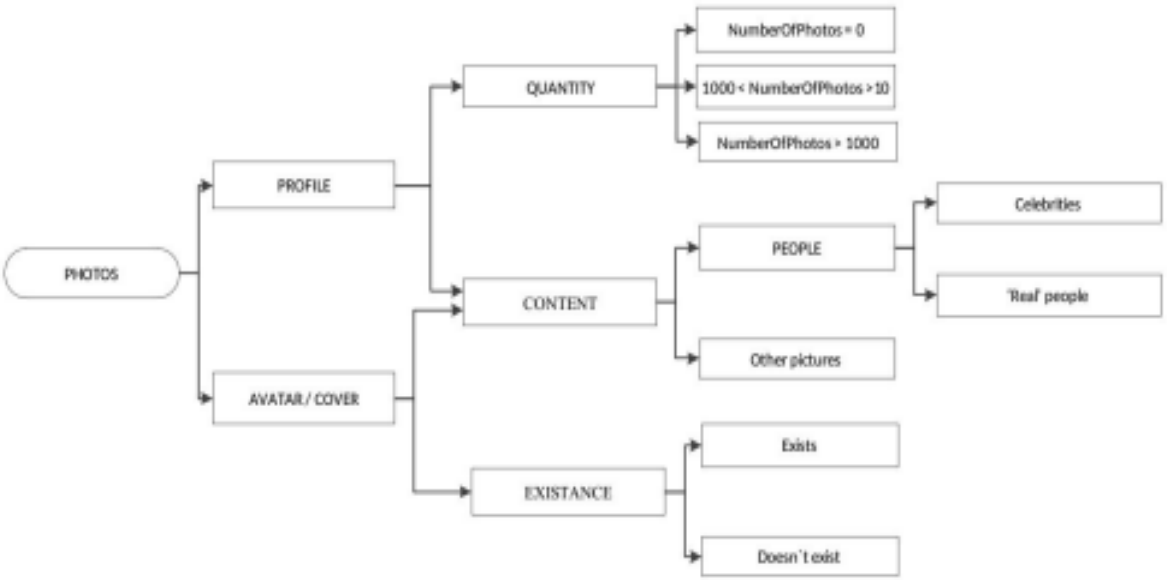

Рис. 3.7. Структурна модель ознак фейковості у категорії «Фото»

Параметри моделі ознак фейковості у категорії «Фото» можна записати у вигляді кортежів:

PHOTO = {PROFILE; AVATAR / COVER}

PROFILE = {QUANTITY; CONTENT}

AVATAR / COVER = {CONTENT; EXISTANCE}

QUANTITY = {NumberOfPhotos}

CONTENT = {PEOPLE; Other pictures}

PEOPLE = {Celebrities; ‘Real’ people}

EXISTANCE = {Exists; Doesn't exist}

Звичайно, окремо ці критерії не можуть однозначно вказувати на «фейковість» облікового запису, оскільки лише аналіз їх об'єднання може поставити під сумнів справжність облікового запису. Для більш точного визначення статусу облікового запису необхідно використовувати аналіз з використанням якомога більшої кількості критеріїв.

У роботі не враховані інші параметри облікових записів, такі як час створення сторінки, швидкість формування кола друзів, зв’язки між друзями користувача, перевірка друзів на фейковість, аналіз фотографій на предмет збігу з іншими зображеннями в інтернеті тощо [31]. Ці та інші параметри будуть враховані у подальших дослідженнях.

3.3 Реалізація системи прийняття рішення

У подальшому дослідженні для прийняття рішення щодо фейковості облікового запису нами обрано метод опорних векторів.

Така класифікація даних методом опорних векторів (Support Vector Machine, SVM) має досить широке застосування [32]. Завдання класифікації полягає у визначенні до якого класу з принаймні двох відомих належить цей об'єкт. Для визначення того, чи є обліковий запис користувача фейковим, визначено два класи, а саме:

«Real»;

«Fake».

Оскільки класів всього два («Fake» / «Real»), то завдання називається бінарною класифікацією. Також в даному випадку існують зразки кожного класу - об'єкти, про які заздалегідь відомо, до якого класу вони належать. Вони знаходяться у відповідному файлі *.csv у вигляді бази даних та в подальшому будуть використані в якості навчальної вибірки.

Отже, математичне формулювання задачі класифікації наступне: нехай *X* - простір об'єктів (*наприклад, P_1, P_2, P_3, .. ., P_{10}*), *Y* - класи (*наприклад, Y = {0, 1}, де 0 - справжній обліковий запис, 1 - фейковий обліковий запис*). Потрібно побудувати функцію *F: X—›Y* (класифікатор), що ставить у відповідність клас *Y*, довільному об'єкту x_i.

Класифікатор працює таким чином: дані точки на площині, що представляють множини параметрів облікових записів, розбиті на два класи (рис. 3.8). Проведено лінію, що розділяє ці два класи. Далі, всі нові точки, які будуть формуватися під час дослідження облікових записів (не з навчальної вибірки), автоматично класифікуються наступним чином: точка вище прямої потрапляє до класу «Fake», точка нижче прямої - до класу «Real». Така пряма називається розділяючою прямою.

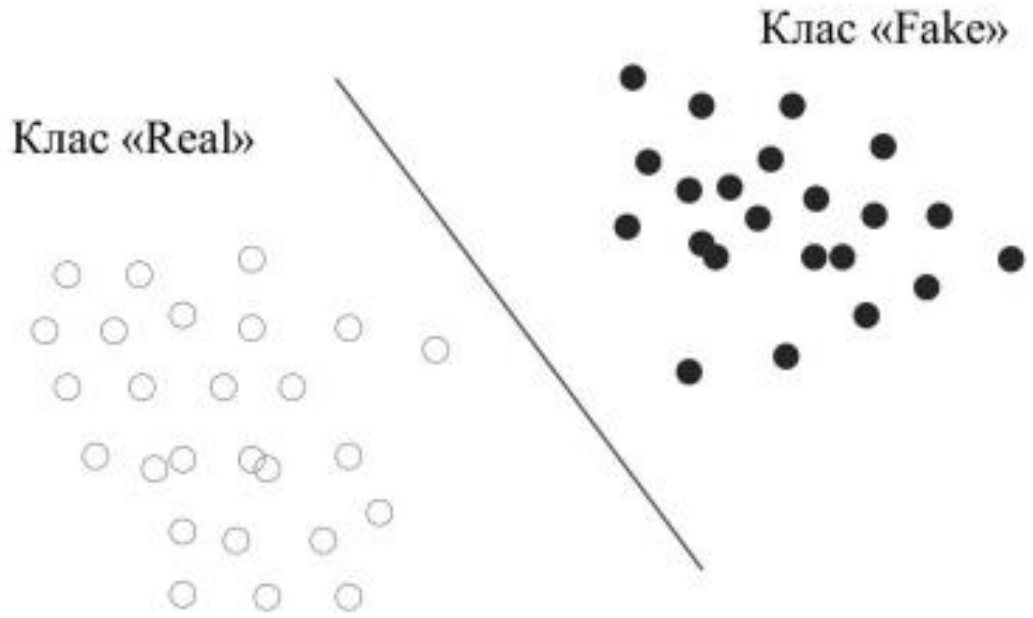

Рис. 3.8. Пряма, що розділяє класи «Real» та «Fake»

Однак, замість прямих доцільно розглядати гіперплощини - площини, розмірність яких на одиницю менша, ніж розмірність початкового простору.

Нехай є навчальна вибірка. Метод опорних векторів будує класифікуючу функцію *F*. Ті об'єкти, для яких *F(X) =1* потрапляють в один клас, а об'єкти з *F(X) = 0* - в інший.

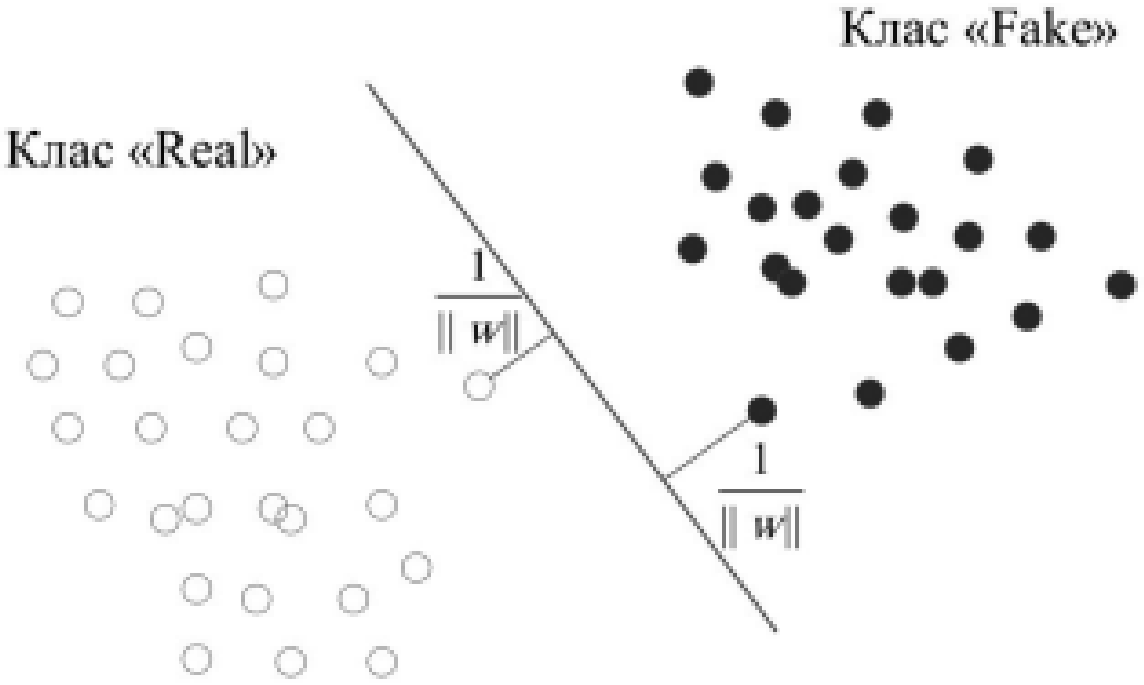

Рис. 3.9. Гіперплощина, що розділяє класи «Real» та «Fake»

Після цього, необхідно вибрати такі *w* і *b,* які максимізують відстань до кожного класу. Дана відстань становить 1/ll w ll (рис. 3.9). Проблема знаходження максимуму 1/ll w ll еквівалентна проблемі знаходження мінімуму ll w ll.

Задача оптимізації мас такий вигляд:

$$\begin{cases} \arg\min\limits_{w,b} \| w\|^2, \\ y_i(\langle w, x_i\rangle + b) \geq 1, i = 1,\ldots,10. \end{cases}$$

(3.1)

Задача оптимізації є стандартною задачею квадратичного програмування і вирішується за допомогою множників Лагранжа. Випадки, коли дані можна розділити гіперплощиною, або, як ще кажуть, лінійно, досить рідкісні. Тому, в цьому випадку необхідно всі елементи навчальної вибірки вкласти у простір *W* більш високої розмірності за допомогою спеціального відображення. При цьому відображення *P* вибирається так, щоб в новому просторі *X* вибірка була лінійно роздільна.

Найчастіше на практиці зустрічаються такі типи ядер: поліноміальні, радіальна базисна функція, гаусова радіальна базисна функція та сигмоїд. Для подальшої розробки нейронної мережі обрано тип ядра сигмоїд.

Нейронна мережа складається з великої кількості однотипних елементів - нейронів, пов'язаних між собою. На рисунку 3.10 зображено схему нейрона.

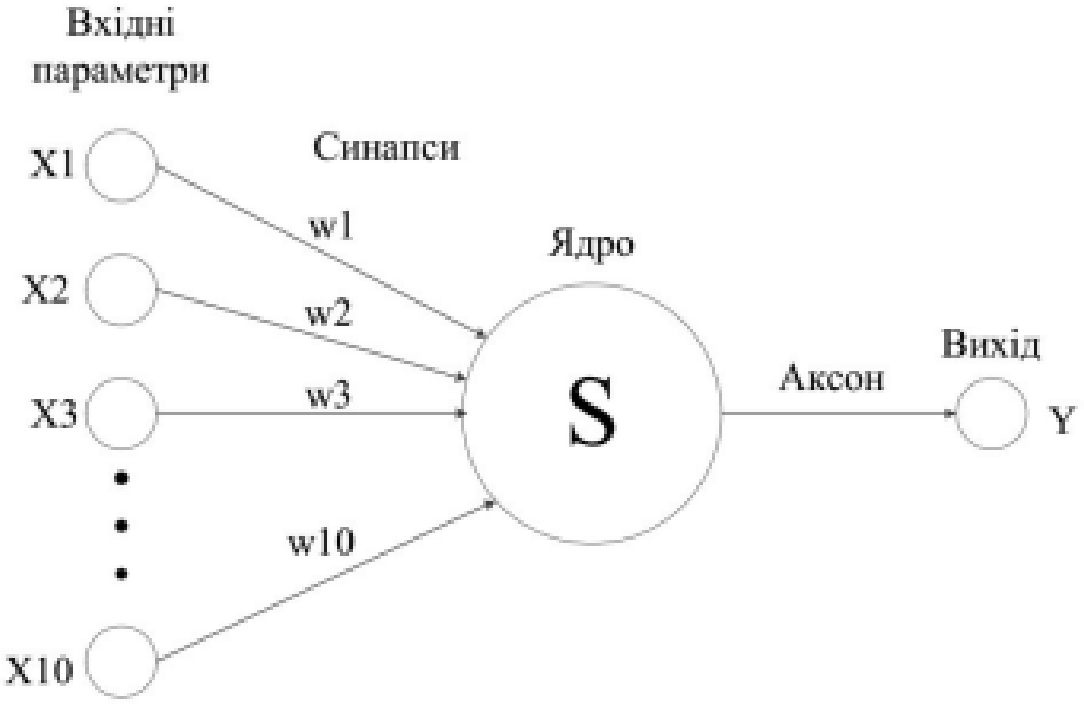

Рис. 3.10. Схема нейрона

Зі схеми зрозуміло, що штучний нейрон складається з таких елементів: синапсів, що пов'язують входи нейрона з ядром; безпосередньо ядра нейрона, що здійснює обробку вхідних даних; аксона, що пов'язує нейрон з наступним шаром нейронів. Кожен синапс має певну вагу, яка визначає вплив відповідного входу нейрона на його стан. Стан нейрона визначається за формулою:

$$S = \sum_{i=1}^{n} x_i w_i \qquad (3.2)$$

Після цього визначається за формулою значення аксона нейрона U= *F(S)*, де *F* - функція, яка називається функцією активації. Найбільш часто в якості функції активації використовується так званий сигмоїд, який мас наступний вигляд:

$$F(x) = \frac{1}{1 + e^{-ax}}$$

(3.3)

У даному випадку задача навчання нейронної мережі зводиться до знаходження функціональної залежності *Y= F(X)*, де *X*— вхідний вектор, а *Y* - вихідний вектор. Вхідний вектор *X* складається з нормалізованого представлення даних про обліковий запис користувача, такі як кількість фотографій, персональна інформація, дата народження тощо.

Основна відмінність цієї нейронної мережі в тому, що в ній всі вхідні і вихідні параметри представлені у вигляді цілих чисел в діапазоні [0, 1] (для вихідних параметрів) та [0, 1, 2] (для вхідних параметрів). У той же час дані предметної області часто кодуються іншим способом. Це можуть бути числові значення у довільному діапазоні, дати, символьні рядки тощо. Таким чином інформація може бути як кількісною так і якісною. Тому необхідно спочатку перетворити якісні дані в числові, а потім перетворити вхідні дані в необхідний діапазон. Таким чином для кожного якісного значення вхідних параметрів (від P_1 до P_{10}) необхідно поставити у відповідність числове значення.

Оптимальними числовими значеннями вхідних параметрів є діапазон від 0 до 2, де 0 - ознака справжності облікового запису, 2 - ознака фейковості, 1 - підозріле значення параметра. Однак, це може спричинити небажану впорядкованість, яка може спотворити дані, і значно ускладнити процес навчання. Проте, у даному випадку для вирішення задачі виявлення фейків такий спосіб перетворення якісних показників у кількісні не модифікує подальші результати та не спотворить результат, що буде обчислено нейронною мережею.

3.4 Розробка програмного засобу

Для роботи з даними соціальної мережі Facebook обрано мову програмування Python, оскільки вона дозволяє обробляти дані великої

розмірності, є простою у використанні, а також дозволяє вирішувати будь-які задачі. Для отримання даних з соціальної мережі «Facebook» обрано модуль Selenium, який дозволяє за допомогою пошуку за елементами веб-сторінок зчитувати з них дані. Для розробки графічного інтерфейсу обрано модуль Tkinter, оскільки він с відносно простим та ефективним під час розробки. Для того, щоб отримати доступ до інформації про користувача у соціальній мережі Facebook, необхідно зчитувати усі дані про користувача разом, що є складним для подальшої обробки інформації. Така складність отримання інформації з соціальної мережі «Facebook» зумовлена нещодавніми подіями масштабного витоку персональної інформації про користувачів, а також часті хакерські атаки на веб-сервери «Facebook». Тому отримання прав розробника для комфортного отримання усіх необхідних даних наразі є досить складною та тривалою процедурою. Для прийняття рішення щодо статусу облікового запису обрано нейронну мережу, яка вирішує задачу класифікації вибору між фейковим та справжнім обліковим записом [10, 29]. Для навчання нейронної мережі створено вибірку даних, яка складається з інформації про існуючі облікові записи, а також їх класифікація.

Архітектура системи програмного засобу для виявлення фейкових облікових записів у соціальній мережі «Facebook» складається з таких компонентів (рис. 3.11):

- власне, програмного засобу, який, у свою чергу, складається з модуля зчитування та обробки інформації, а також нейронної мережі, яка має 9 входів та 1 вихід;
- веб-сторінки соціальної мережі «Facebook», де знаходиться інформація про облікові записи;

•зовнішніх файлів, у яких міститься допоміжна для роботи програми інформація.

Рис. 3.11. Архітектура програмного засобу

Система працює таким чином: після запуску програмного засобу та вибору облікового запису для перевірки відбувається запит до веб-сторінки користувача у «Facebook». У відповідь надходить уся інформація про користувача у текстовому вигляді, що с незручним для обробки. Аналіз даних відбувається за допомогою функцій, що розділяють отриману інформацію та дістають лише необхідну для подальшого аналізу. Програмний засіб перевіряє отриману інформацію та представляє її у вигляді, зрозумілому для системи підтримки прийняття рішень та записує її у файл *account.csv*. Система підтримки прийняття рішень представляє собою нейронну мережу, якій, для достовірності подальших результатів, необхідно навчитися розрізняти справжні та фейкові облікові записи. Для цього використовується навчальний набір даних, що міститься у файлі *training.csv*. Коли система підтримки прийняття рішень готова для обробки даних, на вхід подається файл account.csv з інформацією про обліковий запис, що потребує перевірки. Нейронна мережа обробляє інформацію та надсилає результат обробки головному модулю програми. Після цього програмний засіб виводить на екран результат роботи програмного засобу у вигляді статусу облікового запису та діаграми впливу параметрів на результат.

Розроблені модулі у вигляді класів дозволяють зчитувати та обробляти інформацію про фото на аватарі, фото на фоні обкладинці сторінки, кількість фотографій, кількість друзів, кількість постів, персональна інформація про користувача (місце роботи, освіта, місто проживання/народження, сімейний стан), контактна інформація, дата народження, кількість лайків на аватарі та кількість коментарів на аватарі. Модулі збору інформації працюють паралельно для пришвидшення процесу збору інформації. Таким чином, алгоритм роботи програми можна відобразити у вигляді блок-схеми (рис. 3.12).

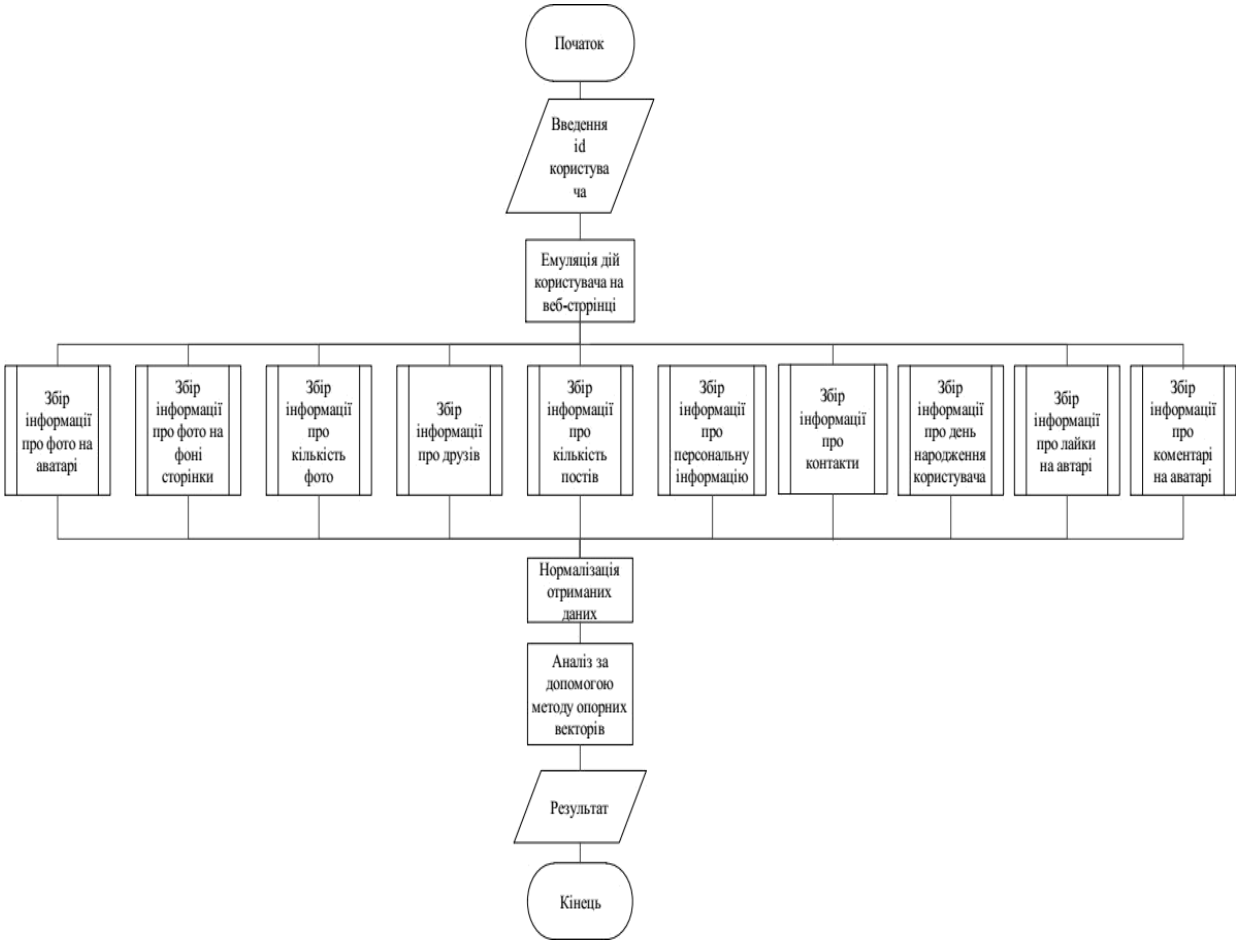

Рис. 3.12. Схема роботи програмного засобу

Як показник обрано систему балів, що свідчить про фейковість облікового запису користувача. Кожен з параметрів під час аналізу залежно від умови отримує певну кількість балів від 0 до 2 (таблиця 3.1).

Таблиця 3.1

Критерії встановлення балів відповідно до значень метрик

Бали / Параметр	0	1	2
Наявність фото на аватарі	На фото зображена людина	Фото існує	Фото не існує
Наявність фото на фоні	Фото існує	-	Фото не існує
Кількість фотографій	Фото не існує	Кількість фото < 10 & Кількість фото > 1000	10 < Кількість фото < 1000
Кількість друзів	10 < Кількість друзів < 2000	0 < Кількість друзів < 10 & Кількість друзів > 2000	Друзів немає
Кількість постів	10 < Кількість постів < 500	Кількість постів > 500 & Кількість постів < 10	Пости не існують
Наявність персональної інформації	Всі поля заповнено	Частина полів заповнено	Інформація відсутня
Контактна інформація	Контакти присутні	-	Контакти відсутні
Дата народження	1932 < Рік народження < 2009	2009 < Рік народження || Рік народження < 1932	Рік народження відсутній
Лайки на аватарі	Кількість лайків	Лайки від друзів < Лайки від чужинців	Лайки відсутні
Кількість коментарів на аватарі	5 < Кількість коментарів < 100	Кількість коментарів < 5 & Кількість коментарів > 100	Коментарі відсутні

Отримані значення для кожного параметру формуються у масив даних, який подається на вхід нейронної мережі, яка, у свою чергу, аналізує їх та видає результат. Залежно від ситуації, існує необхідність проведення додаткових досліджень за участю експертів та з урахуванням інформації, що міститься у гістограмі, для найточнішого визначення статусу облікового запису. Для цього необхідно перевірити кожен зі стовпців та визначити їх рівень впливу. Всього є 3 рівня впливу (від 0 до 2), які вказують, наскільки той чи інший показник вплинув на висновок щодо фейковості облікового запису. Рівні впливу мають такі значення:

0 - не має впливу;

1 - має незначний вплив;

2 - має значний вплив;

Наприклад, якщо стовпець діаграми сягнув 2 рівня, це означає, що параметр, який характеризує даний стовпець, значно вплинув на фейковість облікового запису. У іншому випадку, якщо стовпець знаходиться на 0 рівні, це означає, що параметр відповідає такому, який притаманний справжньому обліковому запису. Таким чином, опираючись на рівні кожного з параметрів, можна зробити висновок про фейковість чи справжність конкретного облікового запису.

Висновки до розділу 3

В розділі проаналізовано можливу інформацію, що міститься в облікових записах у соціальній мережі «Facebook», а також виокремлено основні метрики облікових записів та віднесено їх до відповідних категорій. Розроблено модель системи підтримки прийняття рішень з використанням нейронної мережі, оскільки вона підходить для роботи з нечіткими множинами та великою кількістю даних. Розглянуто структуру нейрона та принцип роботи нейронної мережі. Розроблено алгоритм, архітектуру та схему роботи програмного засобу. Розподілено ступені фейковості облікових записів у вигляді балів відповідно до значень метрик.

ВИСНОВКИ

Розглянуто основні складові фейкових облікових записів, соціальних мереж та використання фейків як для особистої анонімності користувача, так і для ведення інформаційних протиборств. Проаналізовано основні методи класифікації інформації, отриманої з соціальних мереж, а також розглянуто теоретичні відомості про архітектури облікових записів і метрики, які використовують для подальшого аналізу отриманих даних про користувачів, їх групи та інші структури соціальних мереж. Поставлено завдання розробки системи виявлення фейкових облікових записів у соціальних мережах, базуючись на інформації, отриманій з проаналізованих літературних джерел.

В результаті проведеного дослідження структури облікового запису у соціальних мережах виділено фрагменти інформації про користувачів, яка міститься у них. Розглянуто та структуровано метрики за категоріями і значеннями та можливістю їх використання для визначення фейковості облікового запису. Виділено такі категорії: фото, друзі, персональна інформація про користувача, його пости та статуси. Для зручності подальшої обробки всі метрики були віднесені до певних категорій. Розглянуто та проаналізовано існуючі системи підтримки прийняття рішень та обрано для подальшого дослідження бінарний класифікатор, що працює за методом опорних векторів. На основі обраної системи підтримки прийняття рішень реалізовано нейронну мережу для виявлення фейкових облікових записів у соціальній мережі «Facebook».

Розроблено архітектуру та схему роботи програми, а також її основні компоненти. Створено графічний інтерфейс програмного засобу, що є досить зручним для використання. Програмний засіб розроблено за допомогою об'єктно-орієнтованої мови програмування Python, яка дозволяє працювати з даними великої розмірності. Проведено низку експериментальних досліджень для тестування роботи програми шляхом перевірки на фейковість 100 облікових записів у соціальній мережі «Facebook». Експериментальні

дослідження показали достовірність роботи системи виявлення фейкових облікових записів у соціальній мережі «Facebook» на рівні 94%, при цьому помилки першого та другого роду становлять 4% і 2% відповідно.

У подальших дослідженнях планується розширити кількість метрик для аналізу, серед яких - зв’язки між друзями користувача, частота оновлення інформації у обліковому записі, аналіз спільнот та інтересів користувача. Аналіз цих метрик дозволить задіяти більше даних про обліковий запис. Також планується аналізувати облікові записи у інших популярних соціальних мережах, таких як «Twitter» тощо.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Collins 2017 Word of the Year Shortlist. *Collinsdictionary*. URL: https://blog.collinsdictionary.com/language-lovers/collins-2017-word-of-the-year-shortlist/

2. N. Belloir, W. Ouerdane, O. Pastor, É. Frugier, L.-A. de Barmon. A Conceptual Characterization of Fake News: A Positioning Paper. Research Challenges in Information Science. 2022. URL: https://link.springer.com/chapter/10.1007/978-3-031-05760-1_41

3. David Buckingham. Teaching Media in a 'Post-Truth' Age: Fake News, Media Bias and the Challenge for Media Literacy Education. *Cultura y Educación.* 2019. Vol. 31(2). URL: https://davidbuckingham.net/wp-content/uploads/2019/10/post-truth-buckingham.pdf

4. Fallis D., Mathiesen K. Fake news is counterfeit news. *Inquiry*. 2019. P. 1–20. URL: https://doi.org/10.1080/0020174x.2019.1688179

5. Michaelson E., Sterken R., Pepp J. What's New About Fake News?. *Journal of Ethics and Social Philosophy*. 2019. Vol. 16, no. 2. URL: https://doi.org/10.26556/jesp.v16i2.629

6. Tandoc E. C., Lim Z. W., Ling R. Defining "Fake News". *Digital Journalism*. 2017. Vol. 6, no. 2. P. 137–153. URL: https://doi.org/10.1080/21670811.2017.1360143

7. Rahmanian E. Fake news: a classification proposal and a future research agenda. *Spanish Journal of Marketing - ESIC*. 2022. URL: https://doi.org/10.1108/sjme-09-2021-0170

8. C. Wardle. Information disorder: Toward an interdisciplinary framework for research and policy making. 2017. URL: https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html

9. Yadav R. A Simplistic Overview of Machine Learning. *International Journal of Scientific Research in Science, Engineering and Technology*. 2021. P.

184–187. URL: https://doi.org/10.32628/ijsrset218475

10. B. Sharma. An Explanation of Machine Learning. *International Journal of Computer Science and Mobile Computing*. 2019. Vol.8 Issue.4. URL: https://ijcsmc.com/docs/papers/April2019/V8I4201918.pdf

11. Five machine learning types to know. *IBM*. URL: https://www.ibm.com/blog/machine-learning-types/

12. S. Mohamed, R. Ashraf, A. Ghanem, M. Sakr. Supervised Machine Learning Techniques: A Comparison. 2022. URL: https://www.researchgate.net/publication/363870735_Supervised_Machine_Learning_Techniques_A_Comparison

13. Comparing different supervised machine learning algorithms for disease prediction / S. Uddin et al. *BMC Medical Informatics and Decision Making*. 2019. Vol. 19, no. 1. URL: https://doi.org/10.1186/s12911-019-1004-8

14. Self-supervised learning: The dark matter of intelligence. *Meta*. 2021. URL: https://ai.meta.com/blog/self-supervised-learning-the-dark-matter-of-intelligence/

15. C A Padmanabha Reddy Y., Viswanath P., Eswara Reddy B. Semi-supervised learning: a brief review. *International Journal of Engineering & Technology*. 2018. Vol. 7, no. 1.8. P. 81. URL: https://doi.org/10.14419/ijet.v7i1.8.9977

16. B. Dhanalaxm. Machine Learning and its Emergence in the Modern World and its Contribution to Artificial Intelligence. *2020 International Conference for Emerging Technology (INCET)*. 2020. URL: https://ksra.eu/wp-content/uploads/2020/11/10.1109@INCET49848.2020.9154058.pdf

17. Emerging technologies whitepaper series: Machine Learning. *Office of information and communication technology*. 2018. URL: https://unite.un.org/sites/unite.un.org/files/emerging-tech-series-machine-learning.pdf

18. T. R.A. Bakker. Objects in the Age of Virtual Reproduction: Aura and the Elusive Third Axis. 2018. URL: https://core.ac.uk/download/pdf/157730468.pdf

19. R. Dini. An Analysis of Walter Benjamin's The Work of Art in the Age of

Mechanical Reproduction. 2017. URL: https://dokumen.pub/an-analysis-of-walter-benjamins-the-work-of-art-in-the-age-of-mechanical-reproduction-9781912304042-9781912284757-9781912284894.html

20. C. A. Gislam. The Effect of the Non-Human on the Generation of Narrative and Space in Digital Games. *Department of English Manchester Metropolitan University*. 2023. URL: https://e-space.mmu.ac.uk/634136/1/Charlotte_Gislam_PhD.pdf

21. E. Avilés. My/Mi lengua franca: Language, Manipulation, and Cultural Heritage in Chicana Art and Literature. 2014. URL: http://digitalrepository.unm.edu/span_etds/4

22. Manaris B. Natural Language Processing: A Human-Computer Interaction Perspective. *Advances in Computers*. 1998. P. 1–66. URL: https://doi.org/10.1016/s0065-2458(08)60665-8

23. Sawicki J., Ganzha M., Paprzycki M. The state of the art of Natural Language Processing - a systematic automated review of NLP literature using NLP techniques. *Data Intelligence*. 2023. P. 1–47. URL: https://doi.org/10.1162/dint_a_00213

24. Mah P. M., Skalna I., Muzam J. Natural Language Processing and Artificial Intelligence for Enterprise Management in the Era of Industry 4.0. *Applied Sciences*. 2022. Vol. 12, no. 18. P. 9207. URL: https://doi.org/10.3390/app12189207

25. Perifanis N.-A., Kitsios F. Investigating the Influence of Artificial Intelligence on Business Value in the Digital Era of Strategy: A Literature Review. *Information*. 2023. Vol. 14, no. 2. P. 85. URL: https://doi.org/10.3390/info14020085

26. State of Art for Semantic Analysis of Natural Language Processing.D. Hussen Maulud et al. *Qubahan Academic Journal*. 2021. Vol. 1, no. 2. URL: https://doi.org/10.48161/qaj.v1n2a44

27. Chadha N., Gangwar R. C., Bedi R. Current Challenges and Application of Speech Recognition Process using Natural Language Processing: A Survey. *International Journal of Computer Applications*. 2015. Vol. 131, no. 11. P. 28–31. URL: https://doi.org/10.5120/ijca2015907471

28. Exploring the frontiers of deep learning and natural language processing: A comprehensive overview of key challenges and emerging trends.W. Khan et al. *Natural Language Processing Journal*. 2023. P. 100026. URL: https://doi.org/10.1016/j.nlp.2023.100026

29. Speech Recognition Using Deep Neural Networks: A Systematic Review/ A. B. Nassif et al. *IEEE Access*. 2019. Vol. 7. P. 19143–19165. URL: https://doi.org/10.1109/access.2019.2896880

30. D. W. Otter, J. R. Medina, J. K. Kalita. A Survey of the Usages of Deep Learning for Natural Language Processing. *IEEE transactions on neural networks and learning systems*. 2019. Vol. 20, № 10. URL: https://arxiv.org/pdf/1807.10854

31. Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges.S. O. Abioye et al. *Journal of Building Engineering*. 2021. Vol. 44. P. 103299. URL: https://doi.org/10.1016/j.jobe.2021.103299

32. Mahyoob M., Algaraady J., Alrahaili M. Linguistic-Based Detection of Fake News in Social Media. *International Journal of English Linguistics*. 2020. Vol. 11, no. 1. P. 99. URL: https://doi.org/10.5539/ijel.v11n1p99

33. P. K. Verma, P. Agrawal, I. Amorim, R. Prodan. WELFake: Word Embedding Over Linguistic Features for Fake News Detection. *IEEE transactions on computational social systems*. 2021. Vol. 8, № 4. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9395133

34. Identifying Fake News on Social Networks Based on Natural Language Processing: Trends and Challenges.N. R. de Oliveira et al. *Information*. 2021. Vol. 12, no. 1. P. 38. URL: https://doi.org/10.3390/info12010038

35. García-Díaz J. A., Beydoun G., Valencia-García R. Evaluating Transformers and Linguistic Features integration for Author Profiling tasks in Spanish. *Data & Knowledge Engineering*. 2024. P. 102307. URL: https://doi.org/10.1016/j.datak.2024.102307

36. Linguistic Features and Bi-LSTM for Identification of Fake News. A. A. Ali et al. *Electronics*. 2023. Vol. 12, no. 13. P. 2942. URL:

https://doi.org/10.3390/electronics12132942

37. J. Tompkins. Disinformation Detection: A review of linguistic feature selection and classification models in news veracity assessments. 2018. URL: https://arxiv.org/pdf/1910.12073

38. A Comprehensive survey of Fake news in Social Networks: Attributes, Features, and Detection Approaches..M. R. Kondamudia et al. *Journal of King Saud University - Computer and Information Sciences*. 2023. P. 101571. URL: https://doi.org/10.1016/j.jksuci.2023.101571

39. A Comprehensive Study of ChatGPT: Advancements, Limitations, and Ethical Considerations in Natural Language Processing and Cybersecurity.M. Alawida et al. *Information*. 2023. Vol. 14, no. 8. P. 462. URL: https://doi.org/10.3390/info14080462

40. LSTMCNN: A hybrid machine learning model to unmask fake news.D. G. Dev et al. *Heliyon*. 2024. Vol. 10, no. 3. P. e25244. URL: https://doi.org/10.1016/j.heliyon.2024.e25244

41. Nasir J. A., Khan O. S., Varlamis I. Fake news detection: A hybrid CNN-RNN based deep learning approach. *International Journal of Information Management Data Insights*. 2021. Vol. 1, no. 1. P. 100007. URL: https://doi.org/10.1016/j.jjimei.2020.100007

42. Mane S. A. M., Shinde A. StressNet: Hybrid model of LSTM and CNN for stress detection from electroencephalogram signal (EEG). *Results in Control and Optimization*. 2023. P. 100231. URL: https://doi.org/10.1016/j.rico.2023.100231

43. Sustainable Development of Information Dissemination: A Review of Current Fake News Detection Research and Practice.L. Yuan et al. *Systems*. 2023. Vol. 11, no. 9. P. 458. URL: https://doi.org/10.3390/systems11090458 A Comprehensive Review on Fake News Detection With Deep Learning.M. F. Mridha et al. *IEEE Access*. 2021. Vol. 9. P. 156151–156170. URL: https://doi.org/10.1109/access.2021.3129329

44. Taye M. M. Understanding of Machine Learning with Deep Learning: Architectures, Workflow, Applications and Future Directions. *Computers*. 2023. Vol.

12, no. 5. P. 91. URL: https://doi.org/10.3390/computers12050091

45. Fake news detection: a systematic literature review of machine learning algorithms and datasets.H. F. Villela et al. *Journal on Interactive Systems*. 2023. Vol. 14, no. 1. P. 47–58. URL: https://doi.org/10.5753/jis.2023.3020

46. Sentiment Analysis for Fake News Detection.M. A. Alonso et al. Electronics. 2021. Vol. 10, no. 11. P. 1348. URL: https://doi.org/10.3390/electronics10111348

49. «Наклеп» та «образу» знову намагаються повернути в кримінальний кодекс. Центр демократії та верховенства права. 2016. URL: http://cedem.org.ua/news/naklep-ta-obrazu-znovu-namagayutsya-povernuty-vkryminalnyj-kodeks/

50. Бабак А. Як публікувати контент від очевидців подій: поради журналістам. Media Sapiens. – 2016. – URL: http://osvita.mediasapiens.ua/mediaprosvita/how_to/yak_publikuvati_kontent_vid_ochevidtsiv_podiy_poradi_zhurnalistam/

51. Бородянський В. у соціальній мережі Facebook. – 2016. – URL: https://www.facebook.com/borodyanskiy?hc_ref=PAGES_TIMELINE&fref=nf

52. Впровадження медіаосвіти та медіаграмотності в загальноосвітніх школах України : аналітичний звіт за результатами комплексного дослідження 2014–2016 рр. на замовлення Українського медійного проекту («У-Медіа») Інтерньюз Нетворк.2016 URL: https://www.slideshare.net/umedia/2016-66299447.

53. Горська К. О. Медіаконтент цифрової доби: трансформації та функціонування : дис. … докт. наук з соц. комун. : 27.00.01.Горська Катерина Олександрівна ; М-во освіти і науки України, Київ. нац. ун-т ім. Тараса Шевченка, Ін-т журналістики. – Київ, 2016. – 449 с.

54. Гуменюк Л. Практична медіаграмотність. Посібник для бібліотекарів ./ Л. Гуменюк, В. Потапова ; ред.- упоряд. О. Волошенюк. – 2015. – URL: http://aup.com.ua/books/mbm/

55. Дачковська М. ГО «Детектор медіа» створила перший в Україні онлайн-посібник з медіаграмотності для підлітків . Марія Дачковська. Media

Sapiens. – 2016. – URL: http://osvita.mediasapiens.ua/mediaprosvita/kids/go_detektor_media_stvorila_pershiy_v_ukraini_onlaynposibnik_z_mediagramotnosti_dlya_pidlitkiv/

56. Дорожня карта з медіаосвіти і медіаграмотності України. Медіаосвіта та медіаграмотність. – 2016. – URL: http://www.medialiteracy.org.ua/index.php/biblioteka/publikatsii/19- publikatsii/572-dorozhnya-karta-z-mediaosvity-i-mediahramotnostiukrayiny.html.

57. Дослідження на тему «Впровадження медіаосвіти та медіаграмотності в загальноосвітніх школах України» . ГО «Європейська дослідницька асоціація» за підтримки Internews Network. – 2015. – URL: http://osvita.mediasapiens.ua/content/files/me_in_schools__analitycal_report_may_5_2015_f.pdf

58. Журналісти та технологічні компанії запустили спільну ініціативу для просування медіаграмотності . Media Sapiens. – 2017. – URL: http://osvita.mediasapiens.ua/web/online_media/zhurnalisti_ta_tekhnologichni_kompanii_zapustili_spilnu_initsiativu_dlya_prosuvannya_mediagramotnosti/

59. Закон про наклеп. Диктатура чи світова практика?. Корреспондент.net. – 2016. – URL: http://ua.korrespondent.net/ukraine/3671260-zakon-pro-naklep-dyktatura-chysvitova-praktyka

60. Інститут Екології масової інформації. – URL: http://institutes.lnu.edu.ua/mediaeco/

61. Іщенко А. Ю. Медіаосвіта як чинник підвищення якості освіти та засіб протистояння гуманітарної агресії .: аналітична записка . А. Ю. Іщенко Національний інститут стратегічних досліджень : офіційне інтернет-представництво. – URL: http://www.niss.gov.ua/articles/1795/

62. Концепція впровадження медіаосвіти в Україні (нова редакція) . Media Sapiens. – 2016. – URL: http://osvita.mediasapiens.ua/mediaprosvita/mediaosvita/kontseptsiya_vprovadzhennya_mediaosviti_v_ukraini_nova_redaktsiya/

63. Король Д. Інформація − зброя: кому належать українські ЗМІ .:

власники найбільших телеканалів, радіостанцій, газет, журналів та онлайн-видань.Д. Король, Ю. Віннічук, Д. Косенко Insider. – 2015. – URL: http://www.theinsider.ua/infographics/2014/2015_smi/vlasnyky.html

64. Медіакомпетентність фахівця: колект. моногр..Г. В. Онкович, Ю. М. Горун, В. О. Кравчук та ін. ; за ред. Г. В. Онкович ; Нац. акад. пед. наук України, Ін-т вищ. освіти. − К. : Логос, 2013. − 287 с.

47. Медіа-освіта та медіаграмотність : підручник.ред.-упор. В. Ф. Іванов, О. В. Волошенюк; за наук. редакцією В. В. Різуна. – К. : Центр вільної преси, 2012. – 352 с.

65. Медіаосвіта та медіаграмотність : підручник.ред.- упор. : В. Ф. Іванов, О. В. Волошенюк ; за наук. ред. В. В. Різуна. – Київ : Центр вільної преси, 2012. − 352 с.

66. Міщенко, Л. Д. Метод розпізнавання фейкових новин у мережі інтернет на основі обробки природної мови : дис. … д-ра філософії : 123 Комп’ютерна інженерія. – Київ, 2024. – 232 с.

67. Найдьонова Л. Оновлення Концепції медіаосвіти: навіщо було потрібне і які зміни внесені. Media Sapiens. – 2016. – URL:http://osvita.mediasapiens.ua/mediaprosvita/mediaosvita/onovlennya_kontseptsi i_mediaosviti_navischo_bulo_potribne_i_yaki_zmini_vneseni/

68. Осюхіна М. О. Використання контенту користувачів (UGC) в українських телевізійних новинах: види, частота, особливості подачі . Марина Олександрівна Осюхіна. Образ : науковий журнал. за ред. Н. Сидоренко, О. Ткаченко ; Сумський державний університет ; Інститут журналістики КНУ імені Тараса Шевченка. – Суми ; Київ, 2016. – Вип. 4 (22). – С. 69–78

69. Панчук Д. О. Фейки про російсько-українську війну в зарубіжній журналістиці. - Кваліфікаційна робота на здобуття ступеня магістр спеціальності "Журналістика". - Національний авіаційний університет. - Київ, 2023. – 84 с.

70. Степенко, Б. В. Розпізнавання фейкових новин, використовуючи методи штучного інтелекту : магістерська дис. : 122 Комп'ютерні науки .

Степенко Богдан Валентичнович. – Київ, 2022. – 124 с.

71. Стеченко Д. М. Методологія наукових досліджень : підручник.Д. М. Стеченко, О. С. Чмир. – К. : Знання, 2005. – 309 с.

72. Ходаківський Є. Методологія наукових досліджень в парадигмі синергетики : монографія . Є. Ходаківський, В. Данилко, Ю. Цал–Цалко. – Житомир : ЖДТУ, 2009. – 340 с.

73. Череповська Н. Медіа-культура та медіа-освіта учнів ЗОШ : візуальна медіа-культура.Н. Череповська. – Київ : Шк. світ, 2010. – 128 с.

74. Шепелюк В. М. Застосування медіатехнології як мотивація здібностей у студентів до навчання. URL : http://univerua.rv.ua/VNS1/Shepeluk_V.M.pdf