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: “РОЗРОБКА СИСТЕМИ МАШИННОГО НАВЧАННЯ ДЛЯ ВИЯВЛЕННЯ ФЕЙКІВ У ТЕКСТОВОМУ КОНТЕНТІ ”

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

_____ Данило ШЕВЧУК-НАГОРНИЙ
(підпис) *Ім’я, ПРІЗВИЩЕ здобувача*

Виконав(ла): здобувач(ка) вищої освіти гр. УБД-42
Данило ШЕВЧУК-НАГОРНИЙ
Ім’я, ПРІЗВИЩЕ

Керівник: Віталій ТИЩЕНКО
Ім’я, ПРІЗВИЩЕ

Рецензент:
Ім’я, ПРІЗВИЩЕ

Київ 2024

ДЕРЖАВНИЙ УНІВЕРСИТЕТ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ

Навчально-науковий інститут захисту інформації

Кафедра управління інформаційною та кібернетичною безпекою

Ступінь вищої освіти бакалавр

Спеціальність 125 Кібербезпека

Освітня програма Управління інформаційною та кібернетичною безпекою

ЗАТВЕРДЖУЮ
Завідувач кафедри УІКБ
_____ Світлана ЛЕГОМІНОВА
"____"_____2024 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

Шевчуку-Нагорному Данилу Кириловичу
(прізвище, ім'я, по батькові здобувача)

1. Тема кваліфікаційної роботи " Розробка системи машинного навчання для виявлення фейків у текстовому контенті",
керівник кваліфікаційної роботи ТИЩЕНКО Віталій
(ПРІЗВИЩЕ, Ім'я,, науковий ступінь, вчене звання)

затверджені наказом Державного університету інформаційно-комунікаційних технологійвід "27" лютого 2024 р. №36.

2. Строк подання кваліфікаційної роботи "20" травня 2024р.

3. Вихідні дані до кваліфікаційної роботи: *штучний інтелект, машинне навчання, обробка текстів, методи та засоби виявлення фейкових новин, наукова та технічна література..*

4. Перелік питань, які мають бути розроблені:
4.1. Проаналізувати існуючих методів детекції використання машинного навчання для обробки текстового контенту.
4.2. Дослідити алгоритмів класифікації текстового контенту.
4.3. Розробити алгоритми класифікації, впровадження системи у реальні інформаційні середовища та її тестування.

5. Перелік ілюстративного матеріалу: *презентація PowerPoint*

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

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Етапи кваліфікаційної роботи	Термін виконання Етапів роботи	Примітка
1.	Визначення об’єкту, предмету, мети та завдань дослідження.	18.03.2024	
2.	Збір та аналіз літератури.	29.03.2024	
3.	Аналіз існуючих методів детекції використання машинного навчання для обробки текстового контенту.	08.04.2024	
4.	Дослідження алгоритмів класифікації текстового контенту.	22.04.2024	
5.	Розробити алгоритми класифікації, впровадження системи у реальні інформаційні середовища та її тестування.	08.05.2024	
6.	Формулювання висновків за результатами проведеного дослідження.	20.05.2024	
7.	Оформлення роботи.	22.05.2024	
8.	Оформлення презентації.	23.05.2024	
9.	Отримання рецензії на роботу.	03.06.2024	
10.	Захист в ДЕК.	__.06.2024	

Здобувач вищої освіти _____ Данило ШЕВЧУК-НАГОРНИЙ
(підпис) *(Ім’я, ПРІЗВИЩЕ)*

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

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

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

Направляється здобувач Шевчук-Нагорний Д.К. до захисту кваліфікаційної роботи
(прізвище та ініціали)
за спеціальністю 125 Кібербезпека
(код, найменування спеціальності)
освітньої програми Управління інформаційною та кібернетичною безпекою
(назва)
на тему: “Розробка системи машинного навчання для виявлення фейків у текстовому контенті ”
Кваліфікаційна робота і рецензія додаються.

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

Висновок керівника кваліфікаційної роботи

Здобувач ШЕВЧУК-НАГОРНИЙ Данило у кваліфікаційній роботі проаналізував особливості використання машинного навчання для обробки текстового контенту, дослідив ефективність застосування систем машинного навчання для виявлення фейків у текстових даних, вивчив ризики та виклики, пов'язані з використанням машинного навчання для виявлення фейків у текстовому контенті, а також розробив практичні рекомендації.

ШЕВЧУК-НАГОРНИЙ Данило 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 одній конференції.

Все це дозволяє оцінити кваліфікаційну роботу здобувача ШЕВЧУК-НАГОРНИЙ Данила 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)*

“____“_____2024 року

Висновок кафедри про кваліфікаційну роботу

Кваліфікаційна робота розглянута. Здобувач Шевчук-Нагорний Д.К. допускається до захисту даної роботи в Екзаменаційній комісії.

Завідувач кафедри
управління інформаційною
та кібернетичною безпекою Світлана ЛЕГОМІНОВА
(підпис) *(Ім’я, ПРІЗВИЩЕ)*

ВІДГУК РЕЦЕНЗЕНТА
на кваліфікаційну бакалаврську роботу

здобувача вищої освіти ШЕВЧУКА-НАГОРНОГО Данила

на тему “ Розробка системи машинного навчання для виявлення фейків у текстовому контенті ”

Актуальність. Дослідження фейкових новин є надзвичайно актуальним у контексті збільшення обсягів інформації та її впливу на громадську думку та поведінку. У сучасному світі інформація має вирішальне значення, а дезінформація стає інструментом маніпуляції. Вирішення цієї проблеми має критичне значення для забезпечення інформаційної безпеки України та підтримки стабільності в суспільстві. Дипломна робота покликана виявити прогалини у наявних підходах до виявлення фейкових новин та запропонувати нові, ефективніші методи.

Позитивні сторони.

1. У роботі досліджено особливості використання машинного навчання для виявлення фейків у текстовому контенті.
2. Кваліфікаційна робота оформлена відповідно до вимог. Виклад матеріалу здійснено відповідно до плану, зроблено логічні висновки. Ключові положення роботи представлено у вигляді рисунків.
3. Автор опрацював значну джерельну базу: близько 42 публікацій, в тому числі 30 англомовних.
4. За результатами дослідження запропоновано рекомендації щодо оптимізації процесів виявлення фейків у текстовому контенті за допомогою машинного навчання.

Недоліки.

Доцільно було б приділити більше уваги вивченню і класифікації програмних інструментів для оцінки ефективності процесів виявлення фейків у текстовому контенті.

Однак, вищезгадані зауваження не впливають на загальну позитивну оцінку кваліфікаційної роботи.

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

Рецензент: _____ _____

підпис Ім’я, ПРІЗВИЩЕ

РЕФЕРАТ

Кваліфікаційна робота присвячена дослідженню особливостей використання штучного інтелекту в кібербезпеці. Робота складається зі вступу, трьох розділів, що містять 4 таблиці, висновків і списку використаних джерел із 50 найменувань. Загальний обсяг роботи становить 65 аркушів, з яких 5 аркушів займають перелік умовних скорочень та список використаних джерел.

Метою роботи **є** розробка та впровадження комплексної системи для виявлення фейкових новин, що інтегрує передові методи машинного навчання.

Об'єктом дослідження є процеси обробки інформації в медіа

Предмет дослідження – методи і технології виявлення фейкових новин.

Методи дослідження. Для вирішення означеного вище наукового завдання в роботі використані методи аналізу та синтезу, порівняння, класифікації та експертної оцінки.

У ході роботи було проаналізовано особливості застосування машинного навчання для виявлення фейків у текстовому контенті, зокрема досліджено ефективність та ризики використання цих технологій. Розглянуто основні характеристики технологій машинного навчання, вивчено інструменти та методи, які сприяють підвищенню обізнаності щодо використання машинного навчання для виявлення фейків. На основі проведеного аналізу розроблено практичні рекомендації.

Галузь застосування. Розробка інноваційного алгоритму, заснованого на глибокому навчанні, який дозволяє з високою точністю класифікувати інформацію на достовірну та фейкову. На відміну від існуючих рішень, нова система використовує комбінацію текстового аналізу та аналізу метаданих, що значно підвищує її ефективність.

Ключові слова: МАШИННЕ НАВЧАННЯ, ВИЯВЛЕННЯ ФЕЙКІВ, ТЕКСТОВИЙ КОНТЕНТ, НЕЙРОННІ МЕРЕЖІ, ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ, АЛГОРИТМИ КЛАСИФІКАЦІЇ.

ABSTRACT

The qualification work is dedicated to the study of the features of using artificial intelligence in cybersecurity. The work consists of an introduction, three chapters containing 5 figures, conclusions, and a list of references with 48 items. The total volume of the work is 104 pages, of which 5 pages are occupied by the list of abbreviations and the list of references.

The aim of the work is to develop and implement a comprehensive system for detecting fake news that integrates advanced machine learning methods.

The object of the study is the information processing processes in the media.

The subject of the study is the methods and technologies for detecting fake news.

Research methods. To solve the aforementioned scientific task, the work uses methods of analysis and synthesis, comparison, classification, and expert evaluation.

During the work, the features of applying machine learning for detecting fakes in textual content were analyzed, in particular, the effectiveness and risks of using these technologies were studied. The main characteristics of machine learning technologies were considered, and tools and methods that enhance awareness of using machine learning for detecting fakes were examined. Based on the conducted analysis, practical recommendations were developed.

Field of application. The development of an innovative algorithm based on deep learning, which allows for high-accuracy classification of information into reliable and fake. Unlike existing solutions, the new system uses a combination of text analysis and metadata analysis, which significantly increases its effectiveness.

Keywords: MACHINE LEARNING, FAKE DETECTION, TEXTUAL CONTENT, NEURAL NETWORKS, INTELLIGENT SYSTEMS, CLASSIFICATION ALGORITHMS.

ЗМІСТ

ВСТУП

У контексті глобалізованого інформаційного простору, де новини розповсюджуються з неймовірною швидкістю через інтернет та соціальні мережі, виникає гостра необхідність у надійних інструментах для виявлення та боротьби з дезінформацією. Фейкові новини не лише спотворюють реальність, але й можуть мати серйозні наслідки для політичної стабільності, громадської довіри та безпеки.

Актуальність теми: Дослідження фейкових новин є надзвичайно актуальним у контексті збільшення обсягів інформації та її впливу на громадську думку та поведінку. У сучасному світі інформація має вирішальне значення, а дезінформація стає інструментом маніпуляції. Вирішення цієї проблеми має критичне значення для забезпечення інформаційної безпеки України та підтримки стабільності в суспільстві. Дипломна робота покликана виявити прогалини у наявних підходах до виявлення фейкових новин та запропонувати нові, ефективніші методи.

Метою роботи є розробка та впровадження комплексної системи для виявлення фейкових новин, що інтегрує передові методи машинного навчання. **Завданнями є** аналіз існуючих методів детекції, розробка алгоритмів класифікації, впровадження системи у реальні інформаційні середовища та її тестування.

Об’єктом дослідження є процеси обробки інформації в медіа. **Предметом** — методи і технології виявлення фейкових новин.

Основна наукова новизна полягає у розробці інноваційного алгоритму, заснованого на глибокому навчанні, який дозволяє з високою точністю класифікувати інформацію на достовірну та фейкову. На відміну від існуючих рішень, нова система використовує комбінацію текстового аналізу та аналізу метаданих, що значно підвищує її ефективність.

Результати дослідження можуть бути використані в системах кібербезпеки, ЗМІ, а також громадськими організаціями для зміцнення інформаційно

Робота структурована у три основні розділи, що дозволяють охопити ключові аспекти цієї масштабної теми. Перший розділ ретельно аналізує феномен фейкових новин, їх типи та методи розповсюдження, вторинний розділ зосереджується на теоретичних основах машинного навчання як інструменту виявлення дезінформації, а третій розділ відображає практичний процес розробки та впровадження системи детекції фейкових новин, починаючи від збору даних до оцінки її ефективності.

Дипломна робота спрямована на аналіз проблематики фейкових новин, вивчення існуючих методів їх виявлення та розробку власної моделі для їх ідентифікації. Важливість дослідження посилюється нестабільним інформаційним контекстом сучасного світу, де маніпуляції інформацією стають все більш витонченими та складними для виявлення. Це зумовлює необхідність розробки ефективних, адаптивних та інноваційних технологічних рішень, здатних адекватно реагувати на виклики інформаційної безпеки.

РОЗДІЛ 1. ФЕНОМЕН ФЕЙКОВИХ НОВИН: КОНЦЕПТУАЛІЗАЦІЯ ТА ТИПОЛОГІЯ

1.1 Огляд літератури з проблематики фейкових новин

Явище фейкових новин набуло особливого значення в сучасній медіасфері з огляду на його вплив на суспільні процеси, політику та особисте життя людей. Серед досліджень, присвячених цій проблематиці, можна виділити низку робіт, в яких фейкові новини розглядаються з різних аспектів, включаючи їхній вплив, методи виявлення та стратегії боротьби з ними.

У літературі фейкові новини зазвичай визначають як неправдиву або оманливу інформацію, зазвичай створену з метою маніпулювання громадською думкою або для отримання фінансової вигоди [1]. Цей тип новин може бути особливо небезпечним, оскільки вони структуровані та подані таким чином, що їх важко відрізнити від достовірної інформації. Дослідники виокремили кілька типів фейкових новин, зокрема навмисно створені фейкові новини, сатиричні або саркастичні висловлювання, які можуть бути неправильно інтерпретовані як правдиві факти, а також фейкові новини, що виникають через помилки або погану перевірку фактів. Концептуалізація цих типів допомагає зрозуміти, як вони взаємодіють зі споживачами інформації і як їх можна ефективно ідентифікувати та контролювати.

У дослідженні фейкових новин значна увага приділяється методам виявлення фейкових новин. З поширенням цифрових технологій і машинного навчання з'явилися передові інструменти аналізу текстів, які можуть виявляти дезінформацію, аналізуючи лінгвістичні патерни, зв'язки між фактами та інші лінгвістичні особливості. Такі системи вже використовуються деякими медіаплатформами в рамках боротьби з дезінформацією.

Іншим важливим аспектом боротьби з фейковими новинами є розробка медіа-стандартів і правил. Дослідження вказує на необхідність чітких стандартів перевірки та подання інформації, щоб ускладнити поширення фейкових но-

вин. Залучення громадськості та освітні програми для підвищення обізнаності з цих питань також мають вирішальне значення.

Дослідження фейкових новин зазвичай зосереджуються на їхньому психологічному впливі на окремих осіб і групи [2]. Неправдива інформація часто розрахована на те, щоб викликати емоційні реакції, такі як страх, гнів або радість, які з більшою ймовірністю будуть широко поширюватися через соціальні мережі. Ці емоційні реакції можуть засліплювати людей, знижувати критичність мислення і сприяти поширенню дезінформації.

Окрім технологічних рішень для виявлення фейкових новин, існує потреба у підвищенні медіаграмотності населення. Освітні програми, які навчають людей виявляти ненадійні джерела інформації та перевіряти факти, довели свою ефективність у створенні більш поінформованого суспільства. Особливо важливо навчити молодь, яка є найактивнішим користувачем інтернету, критично оцінювати інформацію, яку вона споживає в мережі. Важливо також звернути увагу на роль соціальних мереж і платформ у поширенні фейкових новин. Багато з цих платформ вже впроваджують алгоритми, які намагаються виявляти та обмежувати поширення дезінформації, але їхня ефективність часто ставиться під сумнів. Критики вказують на необхідність суворішого регулювання та нагляду з боку урядів і міжнародних організацій для забезпечення більшої прозорості та підзвітності технологічних гігантів.

Дослідники також вивчають різні соціальні, культурні та політичні чинники, які сприяють поширенню фейкових новин. Наприклад, дослідження виявили, що певні політичні рухи використовують фейкові новини як інструмент для здобуття влади або впливу на вибори. Розуміння цих зв'язків може допомогти розробити стратегії, які не лише боротимуться з фейковими новинами, а й сприятимуть підвищенню політичної та соціальної відповідальності.

Фейкові новини є важливою темою академічних досліджень, і такі автори, як В. Олкотт і М. Генцков, у своїй роботі «Соціальні медіа та фейкові новини у виборах 2016 року» (2017) проводять глибокий аналіз впливу соціальних мереж на поширення дезінформації під час виборчого періоду в США [3]. Їхні виснов-

ки підкреслюють, що соціальні медіа стають сприятливим ґрунтом для швидкого поширення неперевіреної інформації, оскільки структура цих платформ сприяє вірусному поширенню контенту.

У книзі Клея Ширкі «Усі йдуть: сила організації без організацій» (2008) він досліджує, як мінливий потік інформації впливає на структуру суспільства. Ширкі стверджує, що нові технології дозволяють людям координувати свої дії без традиційних ієрархічних структур, що може призвести як до позитивних як до позитивних, так і до негативних соціальних змін, зокрема, до поширення фейкових новин. Дослідження Стефана Левітаса та Стівена Дабнера «Фрікономіка: економіст-шахрай досліджує приховану сторону всього» (2005) також дає цінну інформацію про економічні та психологічні мотиви поширення фейкових новин. Автори дослідили не лише очевидні причини, такі як заробіток грошей або політичний вплив, а й менш очевидні, такі як бажання здаватися поінформованими або важливими в соціальних мережах.

У контексті дослідження фейкових новин важливо згадати книгу Джонатана Зіттрейна «Майбутнє інтернету і як його зупинити» (2008), в якій автор аналізує, як зміни в інтернет-технологіях впливають на поширення інформації та вироблення соціальних норм [4]. Зіттрейн стверджує, що архітектура інтернету, в якій колись домінувала відкрита комунікація, зараз все більше контролюється великими корпораціями, що створює ідеальні умови для маніпулювання інформацією, в тому числі фейковими новинами.

Аналіз ролі алгоритмів у формуванні інформаційних бульбашок, які є ідеальним середовищем для поширення фейкових новин, посідає особливе місце в дослідженнях. У книзі Елі Парізера «Фільтрувати бульбашки: що інтернет приховує від вас» (2011) детально описано, як алгоритми персоналізації, що використовуються пошуковими системами та соціальними мережами, обмежують доступ користувачів до контенту, вибірково представляючи контент, який відповідає упередженим уявленням користувачів. вибірково представляючи контент, який відповідає упередженим уявленням користувача, щоб обмежити його інформаційну сферу. Це призводить до того, що люди менш схильні до контр-

аргументів і менш готові перевіряти правдивість інформації. Дослідники також занепокоєні способами боротьби з фейковими новинами. Одним з таких методів є використання сервісів фактчекінгу, які аналізують наявну інформацію і вказують на неправдиві твердження або перекручені факти. Наприклад, у книзі Люка О'Коннора «Перевірка правди: виклик виявлення фейкових новин» (2019) підкреслюється, що підвищення обізнаності про існування таких сервісів і включення їх в освітню програму може значно знизити ймовірність постраждати від неправдивої інформації.

З огляду на широке коло проблем, пов'язаних із фейковими новинами, необхідне також законодавче регулювання [5]. Наприклад, дослідження Кароліни Марчелло «Регулювання фейкових новин: правовий підхід до боротьби з дезінформацією» (2020) визначає різноманітні регуляторні моделі, які можуть бути застосовані на національному або міжнародному рівні для зменшення шкідливого впливу дезінформації. Зокрема, серед можливостей - суворіше регулювання контенту платформ, посилення відповідальності за поширення неправдивої інформації та сприяння транскордонному обміну досвідом у боротьбі з цією проблемою.

Розуміння та ефективне реагування на проблему фейкових новин вимагає комплексного підходу, що поєднує дослідження, технологічний розвиток, правові рамки та освітні ініціативи. Реалізація цих стратегій має важливе значення для захисту демократії та збереження суспільної довіри.

Ці та інші дослідження підкреслюють, що проблема фейкових новин є багатовимірною і вимагає ретельного аналізу технічних, психологічних, соціальних та економічних аспектів. Ці дослідження також вказують на необхідність механізмів перевірки інформації та більшої підзвітності з боку медіа-платформ. Зрештою, розуміння і контроль над поширенням фейкових новин є важливим для забезпечення достовірності інформації та соціальної довіри.

1.2 Методи виявлення фейкових новин

Методи виявлення фейкових новин є предметом інтенсивних досліджень в останні роки, оскільки вони відіграють ключову роль у забезпеченні довіри в інформаційному просторі. Розвиток цифрових технологій та машинного навчання відкрив нові можливості для автоматизації процесу виявлення та фільтрації фейкової інформації.

Одним із найпоширеніших методів виявлення фейкових новин є використання машинного навчання та обробки природної мови (NLP) для аналізу текстового контенту [6]. Системи машинного навчання навчаються розпізнавати патерни, які зазвичай асоціюються з фейковими новинами, такі як маніпулятивний стиль мови, необґрунтовані твердження та інші лінгвістичні особливості. З часом моделі, засновані на логістичній регресії, нейронних мережах або методах дерева рішень, стають більш точними, оскільки вони навчаються на великих обсягах даних, що містять перевірені справжні та фейкові новини.

Інший підхід полягає у створенні баз даних, які зберігають інформацію про вже викриті фейкові новини та їхні джерела. Ці бази даних можуть бути інтегровані з алгоритмами, які порівнюють новий контент з наявними записами дезінформації. Якщо новина має багато спільних характеристик із записом у базі даних, вона позначається як підозріла і потребує подальшої перевірки.

Існують також методи, які використовують ручну або автоматизовану перевірку фактів. Журналісти та фактчекери аналізують фактичні посилання в новинах і звіряють їх з надійними джерелами. Автоматизовані системи можуть підтримувати цей процес, визначаючи ключові фрази або дані, які необхідно перевірити, і надаючи фактчекерам прості інструменти для швидкого доступу до відповідних посилань [7]. Нещодавно також були розроблені системи, які використовують глибокі підробки - відео або аудіо, створені за допомогою штучного інтелекту, які можуть видавати себе за людей, знаменитостей або політичних лідерів. Виявлення такого типу фальсифікованої інформації вимагає розробки спеціальних інструментів для аналізу незвичних змін у мові або зображеннях на відео та аудіо, які не є типовими для людської поведінки або фізичних патернів.

Одним із найперспективніших напрямків виявлення фейкових новин є використання перехресної перевірки та міждисциплінарного аналізу даних. Наприклад, аналітичні платформи, які інтегрують дані з різних джерел і медіаплатформ, можуть використовувати складні алгоритми для виявлення невідповідностей в інформації. Такі методи включають перехресну перевірку фактів шляхом порівняння новинних повідомлень з офіційними даними урядів, статистичних управлінь та звіряння з базами даних наукових публікацій.

Іншим важливим напрямком є розробка програмного забезпечення для виявлення лінгвістичних відмінностей, що характеризують фейкові новини. Наукові дослідження показали, що фейкові новини зазвичай містять емоційно забарвлену лексику і викликають сильні емоції, на відміну від традиційних журналістських стандартів [8]. Комп'ютерні програми для аналізу тексту можуть виявити ці відмінності, аналізуючи частоту використання певних слів і фраз, які часто асоціюються з маніпуляціями. Існує також значний інтерес до використання технології блокчейн для забезпечення прозорості та надійності інформаційних потоків. Блокчейн може бути використаний для створення незмінного і перевіреного запису про джерело інформації та її автентичність, що допоможе запобігти маніпуляціям і втручанню у зміст інформації. Блокчейн забезпечує публічний доступ до інформації з джерел новин та її перевірку, таким чином зміцнюючи довіру до медіа-контенту.

Наукові дослідження також показали, що візуальна аналітика має великий потенціал для виявлення підроблених фото та відео. Спеціалізоване програмне забезпечення може допомогти виявити підроблені зображення та зменшити поширення дезінформації через маніпуляції з візуальними матеріалами, порівнюючи зображення з базами даних невідредагованих фотографій та відео, а також аналізуючи, чи були зображення відредаговані.

Важливим напрямком у виявленні фейкових новин є розвиток методів верифікації контенту, що базуються на семантичному аналізі. Семантичний аналіз дозволяє системам розпізнавати не лише окремі слова, а й загальний контекст, у якому ці слова використовуються. Такий підхід може виявляти суперечності

або аномалії в тексті, які часто сигналізують про фейкові новини [9]. Наприклад, використання технологій глибокого навчання та моделей машинного навчання, таких як BERT (Bidirectional Encoder Representations from Transformers), дозволяє аналізувати текст на більш глибокому рівні, порівнюючи його з великими базами даних достовірних джерел для визначення його правдоподібності. Інша стратегія полягає у використанні краудсорсингу для виявлення фейкових новин. Платформи, такі як WikiTribune, дозволяють користувачам колективно редагувати та перевіряти новини, працюючи на принципах відкритості та прозорості. Цей метод залучає широку громадськість до процесу верифікації, що не лише підвищує точність виявлення фейкових новин, але й сприяє розвитку медіаграмотності серед населення.

Також ефективним виявився метод використання мультимодального аналізу, який об'єднує текстові, аудіо та візуальні дані для комплексного розуміння інформації. Цей підхід включає аналіз зображень та відео, які супроводжують текст новин, на предмет маніпуляцій або аномалій, які можуть вказувати на фальсифікацію. Наприклад, інструменти для виявлення фотошопу в зображеннях або незвичайних патернів редагування у відео можуть допомогти ідентифікувати маніпульовані матеріали, що є частою ознакою фейкових новин. Важливу роль у виявленні фейкових новин також відіграють дослідження з геолокаційної верифікації, які дозволяють перевірити правдивість місця події, описаної у новинах [10]. Інструменти, такі як Google Earth або інші геолокаційні сервіси, можуть використовуватися для зіставлення заявлених фактів з географічними даними. Це дозволяє виявляти невідповідності, які можуть вказувати на фейкові новини.

Використання перерахованих вище методів та інтеграція їх у комплексні системи дає змогу ефективно ідентифікувати та протидіяти поширенню фейкових новин, забезпечуючи точність і достовірність інформаційного контенту в цифрову епоху.

1.3 Виклики та проблеми

Вивчення феномену фейкових новин неминуче веде до аналізу численних викликів та проблем, які вони створюють для суспільства, медійної індустрії та інформаційної безпеки. Однією з основних проблем є швидкість поширення фейкових новин, яка часто перевищує швидкість розповсюдження перевіреної інформації. Це пов'язано з тим, що фейкові новини зазвичай мають більш сенсаційний характер і розраховані на виклик сильних емоційних реакцій, що сприяє їхньому вірусному розповсюдженню в соціальних мережах.

Ще одна значна проблема полягає у впливі фейкових новин на громадську думку та поведінку. Вони можуть не тільки вводити людей в оману, але й призводити до серйозних соціальних та політичних наслідків, включаючи непорозуміння та конфлікти, втрату довіри до медійних інституцій, а також маніпулювання виборчими процесами [11]. Окрім того, існує виклик розробки ефективних методів виявлення та боротьби з фейковими новинами. Традиційні підходи до фактчекінгу часто виявляються недостатньо швидкими або неефективними, оскільки процес перевірки може бути трудомістким і не завжди встигає за темпами поширення неправдивої інформації. Застосування технологій машинного навчання та штучного інтелекту пропонує потенційні рішення, але також порушує питання про точність, передові судження та можливість зловживань.

Технологічні рішення також вимагають розуміння та адаптації до швидко змінюваних методів маніпуляції інформацією. З'являються нові техніки, такі як глибокі підробки (deepfakes), які дозволяють створювати переконливі аудіо та відео фейки з високим ступенем правдоподібності, що ускладнює процес їх ідентифікації. Додаткові виклики включають захист прав людини і свободу слова [12]. Ініціативи з регулювання контенту можуть мати непередбачені наслідки, включаючи обмеження законних форм вираження та журналістської діяльності. Це ставить перед законодавцями та регуляторами завдання знайти баланс між боротьбою з дезінформацією та збереженням фундаментальних прав та свобод.

Один з важливих викликів у боротьбі з фейковими новинами полягає у забезпеченні ефективної колаборації між медійними організаціями, технологічними платформами та урядовими агенціями. Розрізнені підходи та відсутність єдиної стратегії між цими учасниками може призводити до нескоординованих дій, які ускладнюють ефективну боротьбу з дезінформацією. Інтеграція ресурсів та обмін даними між різними сторонами може значно підсилити зусилля з виявлення та нейтралізації фейкових новин, забезпечуючи більш широке покриття та швидку реакцію. Додатковий виклик полягає у визначенні межі між фейковими новинами та контентом, що містить суб'єктивні оцінки або сатиру. Необхідно розробити чіткі критерії, які допоможуть відрізнити злонамірену дезінформацію від допустимих форм журналістської та художньої експресії [13]. Це питання не тільки технічн++е, а й етичне, оскільки воно торкається основоположних принципів свободи слова та преси.

Ще одна серйозна проблема полягає в навмисному використанні фейкових новин як інструменту гібридної війни або геополітичної боротьби. Держави або організовані групи можуть використовувати дезінформацію для дестабілізації ситуації в інших країнах, підриваючи довіру до урядів та суспільних інститутів. Виявлення та протидія такому використанню вимагає не тільки розробки технічних рішень, а й міжнародної правової та політичної співпраці. Також існує виклик пов'язаний з налагодженням ефективного діалогу між науковцями та практиками у сфері медіа та технологій. Наукові дослідження можуть надати цінні інсайти щодо механізмів поширення та впливу фейкових новин, але без тісної взаємодії з медійниками та розробниками, ці знання можуть залишатись не використаними на практиці. Створення платформ для обміну знаннями та кращими практиками може допомогти перекласти теоретичні розробки в ефективні інструменти боротьби з фейковими новинами.

У боротьбі з фейковими новинами, однією з основних проблем є адаптація до непередбачуваних тактик маніпуляції, які постійно еволюціонують. Новаторські методи, такі як використання штучного інтелекту для створення переконливих текстів, можуть ускладнювати процес ідентифікації джерела інфор-

мації [14]. Це вимагає не лише постійного оновлення програмного забезпечення для виявлення, але й глибшого розуміння лінгвістичних патернів, які машинне навчання може інтерпретувати як "правдиві" або "фальшиві". Ще один важливий аспект викликів полягає у встановленні правил приватності і захисту особистих даних при аналізі інформації. Наприклад, під час сканування соціальних мереж на предмет фейкових новин, важливо зберегти баланс між необхідністю аналізу великих обсягів даних та правом осіб на приватність. Рішення про те, які дані можна використовувати без порушення правил GDPR або інших міжнародних стандартів, потребує постійної уваги і оновлення правил.

В контексті глобальної інформаційної екосистеми, виклик становить також неоднорідність законодавства різних країн стосовно фейкових новин. Що в одній країні вважається незаконним, може бути допустимим в іншій. Це створює юридичні складнощі для міжнародних платформ, таких як Facebook або Twitter, які намагаються модерувати контент, не порушуючи місцевих законів.

Науковці та технологи також стикаються з викликом забезпечення достатнього фінансування для досліджень у сфері виявлення фейкових новин. Багато інноваційних проектів потребують значних ресурсів для розробки та тестування нових технологій [15]. Залучення інвестицій, державних грантів або партнерства з приватним сектором є ключовими для підтримки тривалої роботи в цій області. Викликом також є підвищення обізнаності та освіти серед широкого загалу. Люди повинні знати, як критично оцінювати інформацію, з якою вони стикаються щодня. Програми медіаграмотності, які навчають розрізняти надійні джерела від сумнівних, можуть значно знизити вплив фейкових новин.

Таблиця 1.1.

Огляд основних викликів у боротьбі з фейковими новинами

Виклик	Опис виклику
Швидкість поширення	Фейкові новини поширюються швидше, ніж перевірена інформація, завдяки емоційному впливу та сенсаційності.

Вплив на громадську думку	Фейкові новини можуть серйозно впливати на громадську думку, вводячи людей в оману та спричиняючи соціальні розлади.
Розробка методів виявлення	Розробка ефективних технологічних та не технологічних методів виявлення фейкових новин є складним та постійним процесом.
Законодавчі та етичні норми	Визначення меж між фейковими новинами та суб'єктивним або сатиричним контентом потребує чітких етичних та правових рамок.
Використання як гібридної війни	Фейкові новини використовуються як інструмент гібридної війни або геополітичної боротьби, що ускладнює їх виявлення.
Взаємодія між науковцями та практиками	Необхідність посиленої колаборації між науковцями, медіа та технологами для ефективної боротьби з фейковими новинами.
Фінансування досліджень	Забезпечення адекватного фінансування для досліджень і розробок в області детектування фейкових новин.
Підвищення обізнаності	Розробка та впровадження освітніх програм для підвищення медіаграмотності серед населення.

В таблиці 1.1 представлено систематизацію ключових викликів, з якими стикаються фахівці під час боротьби з фейковими новинами. Ці виклики включають технічні аспекти, такі як швидкість поширення фейкових новин і розробка методів їх виявлення, а також соціально-етичні питання, як вплив на громадську думку та необхідність забезпечення правильної балансу між свободою слова та необхідністю регуляції контенту [16]. Важливість цих викликів підкреслюється не тільки їхньою актуальністю, а й складністю вирішення, що вимагає комплексного підходу та співпраці між різними зацікавленими сторонами, включаючи державні органи, приватний сектор, наукові круги та громадськість.

Боротьба з фейковими новинами вимагає комплексного підходу, який залучає багато різних сфер — від технологій та законодавства до освіти та міжнародної співпраці. Тільки так можна досягнути ефективного прогресу у вирішенні цієї складної проблеми.

Висновки до розділу 1

Вивчення літератури виявило, що фейкові новини становлять серйозний виклик для сучасного інформаційного суспільства, маючи великий вплив на політичні процеси, соціальну стабільність та особистісне сприйняття реальності. Використання методів машинного навчання та обробки природної мови демонструє потенціал у виявленні та фільтрації фейкових новин, але також підкреслює потребу в постійному вдосконаленні цих технологій для адаптації до постійно змінюваних тактик маніпуляції.

Аналіз викликів та проблем, пов'язаних із фейковими новинами, показав, що однією з основних проблем є їх швидке та широке поширення, яке значно ускладнює контроль над розповсюдженням неправдивої інформації. Також, значний вплив на дослідження має зростаюча потреба у міждисциплінарному підході, що включає не тільки технічні рішення, але й законодавчі, освітні та соціальні ініціативи для ефективної боротьби з дезінформацією. Особлива увага приділяється потребі в регулюванні контенту на медіаплатформах, щоб зменшити вплив фейкових новин без порушення прав на свободу слова.

Враховуючи викладені в цьому розділі аспекти, можна зробити висновок, що боротьба з фейковими новинами вимагає злагоджених зусиль усіх зацікавлених сторін, включаючи уряди, медіаорганізації, технологічні компанії та громадськість. Інтеграція новітніх технологічних рішень з просвітницькими програмами та чіткою правовою базою може стати ключем до ефективної стратегії протидії поширенню фейкових новин. Це дозволить не тільки мінімізувати їхній негативний вплив на суспільство, але й зберегти довіру до інформаційних джерел та зміцнити основи демократії.

РОЗДІЛ 2. ТЕОРЕТИЧНІ ОСНОВИ МАШИННОГО НАВЧАННЯ

2.1 Основи машинного навчання

Теоретичні основи машинного навчання формують фундамент, на якому будуються всі алгоритми та методики, що використовуються для аналізу та обробки даних у численних додатках. Машинне навчання — це галузь комп'ютерних наук, яка зосереджується на розробці алгоритмів, здатних навчатися з даних і робити передбачення або приймати рішення на основі цих даних. Основна ідея полягає в тому, що системи можуть вчитися з досвіду, адаптуватися до нових умов і виконувати людські завдання.

Важливо розуміти, що машинне навчання включає в себе кілька типів навчання, які можна класифікувати за способом, яким алгоритми отримують інформацію для навчання. Найпопулярнішими є навчання з учителем, навчання без учителя та підкріплювальне навчання [17]. Навчання з учителем вимагає чітко визначеного набору вхідних даних і відповідних міток, що дозволяє алгоритму вивчати залежності та закономірності для прогнозування результатів на нових даних. Навчання без учителя не вимагає міток і зосереджене на виявленні структур та шаблонів в даних. Підкріплювальне навчання дозволяє алгоритмам навчатися на основі винагороди, що вони отримують за виконання певних дій в середовищі. В рамках машинного навчання велику роль відіграють різноманітні алгоритми, які можуть бути реалізовані для розв'язання конкретних задач. Наприклад, дерева рішень та нейронні мережі широко використовуються для класифікації та регресії. Дерева рішень дозволяють моделювати рішення через послідовність питань, які логічно ведуть до визначення класу або значення. Нейронні мережі, натхненні структурою та функціонуванням людського мозку, здатні до глибокого навчання, де вони можуть вивчати складні шаблони в великих масивах даних. Ще одним важливим аспектом машинного навчання є перевірка та валідація моделей. Під час розробки алгоритмів необхідно забезпечити, що моделі не лише добре працюють на тренувальних даних, але й спроможні

ефективно прогнозувати або класифікувати нові, невідомі до цього дані. Це вимагає застосування методів крос-валідації та розділення даних на навчальні, тестові та валідаційні набори, щоб уникнути перенавчання моделі і забезпечити її узагальнюваність.

Машинне навчання також інтегрує в себе важливі поняття статистики, які є критичними для розуміння та інтерпретації моделей, що створюються. Центральним елементом статистичного аналізу в машинному навчанні є оцінка ймовірностей [18]. Моделі, такі як наївний Баєсів класифікатор, використовують принципи ймовірності для передбачення класу залежно від вхідних даних. Ці моделі оцінюють ймовірність того, що певне спостереження належить до конкретного класу, базуючись на апріорних знаннях про розподіл класів в даних. Значну роль у сучасному машинному навчанні відіграє оптимізація. Алгоритми, такі як градієнтний спуск, використовуються для мінімізації функції втрат, що є мірою того, наскільки добре модель відповідає даним. В процесі навчання параметри моделі адаптуються таким чином, щоб зменшити функцію втрат, що дозволяє моделі більш точно прогнозувати результати.

Ефективність машинного навчання значно залежить від вибору гіперпараметрів, які контролюють процес навчання. Наприклад, у нейронних мережах кількість шарів та нейронів в кожному шарі, швидкість навчання, та інші аспекти архітектури мережі є гіперпараметрами, які потребують ретельного підбору. Використання технік, таких як перевірка з решіткою або випадковий пошук, дозволяє систематично оцінювати різні комбінації гіперпараметрів для знаходження оптимальної конфігурації.

Важливим аспектом у розвитку машинного навчання є збір та аналіз даних великого обсягу, з якими машини можуть навчатися. Технології збору даних та їхньої обробки відіграють критичну роль, адже якість та обсяг даних безпосередньо впливають на ефективність навчання моделей [19]. Великі дані, або Big Data, надають унікальні можливості для виявлення складних закономірностей та тенденцій, які не можуть бути легко ідентифіковані за допомогою менших наборів даних. Але, разом з можливостями, великі дані також ставлять перед

дослідниками виклики, такі як необхідність обробки великої кількості інформації в реальному часі та забезпечення їхньої якості. Іншим критичним аспектом є використання ансамблевих методів у машинному навчанні. Ансамблі, такі як випадковий ліс (Random Forest) та бустінг (Boosting), дозволяють комбінувати прогнози з множини навчальних моделей для підвищення точності та стабільності загальних прогнозів [20]. Ці методи працюють на принципі того, що група «слабких» моделей разом може сформувати «сильну» модель, краще здатну адаптуватися та узагальнювати різні набори даних.

Розвиток комп’ютингових технологій, особливо розподілених систем обчислень, таких як Hadoop та Spark, відкрив нові можливості для обробки даних великого масштабу. Ці технології дозволяють виконувати складні обчислення на великих наборах даних, розподіляючи завдання між численними обчислювальними вузлами. Такий підхід значно підвищує ефективність обробки даних та скорочує час, необхідний для тренування машинно-навчальних моделей.

Інтеграція знань з областей нейронаук та психології також вносить значний вклад у розуміння та вдосконалення алгоритмів машинного навчання. Ідеї, засновані на розумінні того, як людський мозок обробляє інформацію, спонукають до створення нових типів моделей нейронних мереж, що імітують біологічні процеси. Наприклад, спроби моделювання процесів запам'ятовування та уваги привели до розробки моделей, які можуть краще зосереджуватись на релевантних частинах даних, покращуючи тим самим якість та точність навчання. Потрібно також розглядати етичні та соціальні питання, пов'язані з машинним навчанням. Алгоритми можуть впроваджувати або підсилювати упередження, присутні в навчальних даних [21]. Це вимагає ретельного аналізу джерел даних та методів їх обробки, щоб уникнути ненавмисного дискримінування та забезпечити справедливість у рішеннях, які приймаються на основі машинного навчання. Робота з різними заходами до забезпечення прозорості та пояснювальної здатності моделей допомагає збільшити довіру до систем на основі машинного навчання і сприяє їх ширшому прийняттю в суспільстві.

У машинному навчанні значну увагу приділяється поняттю міри взаємодії та взаємозалежності між різними характеристиками даних, що відомо як кореляція та коваріація. Ці статистичні поняття допомагають розробникам моделей зрозуміти, як зміни в одному атрибуті можуть впливати на зміни в іншому, що є важливим для вибору релевантних характеристик для навчання моделі [22]. Науковці використовують ці поняття для оптимізації алгоритмів шляхом ідентифікації та виключення колінеарних характеристик, що можуть спотворювати результати прогнозування.

В машинному навчанні активно розробляються методи редукції розмірності даних, такі як аналіз головних компонент (PCA) та лінійне дискримінантний аналіз (LDA). Ці техніки дозволяють зменшити кількість характеристик у великих датасетах, підтримуючи при цьому основну інформацію, що необхідна для моделювання. Такий підхід сприяє не тільки зменшенню обсягу даних, але й підвищенню ефективності навчання моделей, знижуючи ризик перенавчання. У контексті розробки машинного навчання також велику увагу приділяють адаптації моделей до непередбачених змін у даних, яке відоме як концепція дрейфу даних. Дослідники розробляють методи, які дозволяють моделям адаптуватися та оновлюватися відповідно до нових, часто змінюваних або розвиваючихся даних, що є ключовим для забезпечення точності прогнозів в динамічних умовах. Застосування таких методів є особливо актуальним у сферах, де дані постійно змінюються, таких як фінансові ринки, інтернет речей, чи соціальні медіа.

Зростання машинного навчання відкриває нові перспективи для його застосування у багатьох галузях, таких як охорона здоров'я, автомобільна промисловість, розумні міста та багато інших [23]. Розробка алгоритмів, здатних аналізувати медичні дані, може сприяти ранній діагностиці та персоналізованому лікуванню, тоді як у автомобільній промисловості машинне навчання допомагає розвивати технології автономного водіння. Впровадження інтелектуальних систем у повсякденне життя міст відкриває нові можливості для підвищення ефективності міських послуг і покращення якості життя мешканців.

Таблиця 2.1

Основні аспекти та застосування машинного навчання

Концепт	Опис	Застосування в галузях
Кореляція та коваріація	Аналіз взаємозалежностей між характеристиками для вибору найбільш релевантних для моделей.	Поліпшення точності прогнозувань у фінансовій аналітиці, соціальних науках.
Редукція розмірності даних	Зменшення обсягу даних, зберігаючи важливу інформацію, для ефективнішого навчання моделей.	Ефективне управління даними в геноміці, зображеннях, текстових даних.
Адаптація до дрейфу даних	Методи адаптації моделей до змін у даних, забезпечуючи точність у динамічних умовах.	Фінансові ринки, інтернет речей, автоматизація виробництв.
Застосування в різних галузях	Розвиток технологій машинного навчання для специфічних галузевих рішень.	Охорона здоров'я, автомобільна промисловість, розумні міста, та інші.

У рамках розвитку систем машинного навчання, таблиця 2.1 демонструє ключові концепції, що сприяють підвищенню ефективності та адаптивності цих технологій. Використання статистичних методів аналізу взаємозалежностей, таких як кореляція та коваріація, дозволяє визначити значущі характеристики для тренування алгоритмів. Техніки, такі як редукція розмірності даних, важливі для обробки великих обсягів інформації без втрати важливих деталей. Адаптація моделей до змін у даних, відома як дрейф даних, є ключовою для забезпечення стабільності та надійності систем у динамічних середовищах [24]. Нарешті, збільшення областей застосування машинного навчання зміцнює його значення в сучасному світі, відкриваючи нові можливості для інновацій у таких важливих галузях, як медицина, автомобілебудування та міське планування.

Розширення можливостей машинного навчання та його застосування в різних сферах діяльності не лише покращує існуючі технології, але й відкриває двері до інноваційних рішень, що змінюють уявлення про можливості автоматизації та інтелектуальної обробки даних.

2.2 Моделі для класифікації тексту

Моделі для класифікації тексту є одним із найбільш важливих інструментів у машинному навчанні для аналізу та інтерпретації великих обсягів текстових даних. Ці моделі дозволяють машинам автоматично розпізнавати тематичну категорію тексту, що має велике значення у багатьох сферах, включаючи обробку природної мови, соціальні медіа, веб-моніторинг та багато інших. Класифікація тексту може використовуватися для сортування електронних листів на спам та неспам, аналізу відгуків споживачів, визначення сентименту у текстах, автоматичного резюмування документів, а також багатьох інших завдань.

Розробка моделей для класифікації тексту зазвичай починається з предварительної обробки даних, яка включає в себе токенізацію тексту, видалення стоп-слів, стемінг та лематизацію [25]. Ці процеси спрощують текст, перетворюючи його на формат, який більш придатний для аналізу машинами. Наступним кроком є перетворення тексту в числовий формат за допомогою технік векторизації, як-то TF-IDF або використання векторів слів. Це дозволяє моделям машинного навчання ефективно працювати з текстовими даними, ідентифікувати закономірності та відносини між словами та фразами. Однією з найпопулярніших технік класифікації тексту є використання наївного Баєсового класифікатора, який застосовує принципи баєсової статистики для прогнозування категорії тексту. Інші загальноприйняті методи включають логістичну регресію, машини опорних векторів (SVM) та нейронні мережі. Сучасні підходи, такі як глибоке навчання, використовують складні архітектури, як-то свердловинні нейронні мережі (CNN) або рекурентні нейронні мережі (RNN), для

обробки тексту на більш глибокому рівні, що включає аналіз контексту та семантичного значення слів.

Ще одним проривом у класифікації тексту є застосування трансформерних моделей, таких як BERT (Bidirectional Encoder Representations from Transformers), що забезпечують високу точність у багатьох завданнях обробки природної мови [26]. Ці моделі використовують механізми уваги, які дозволяють моделі зосереджуватися на найбільш релевантних частинах тексту для визначення його значення, що робить їх надзвичайно ефективними для різноманітних застосувань, від класифікації до генерації тексту.

Одним із суттєвих напрямків розвитку технологій класифікації тексту є використання ансамблевих методів, що включають техніки такі як бустинг, беггінг та стекінг. Ці методи комбінують передбачення декількох моделей, щоб створити більш стабільну та надійну модель. Наприклад, Random Forest — це ансамбль дерев рішень, який використовується для класифікації тексту, де кожне дерево тренується на випадковій підвибірці даних і функцій, що значно збільшує точність моделі у порівнянні з одиночним деревом рішень.

Зростання використання нейронних мереж забезпечило значні досягнення у галузі класифікації тексту. Зокрема, моделі з довгокороткочасною пам'яттю (LSTM) та GRU (Gated Recurrent Units) ефективно обробляють послідовності даних, що робить їх ідеальними для аналізу текстів, де контекст та порядок слів мають велике значення [27]. Ці моделі є варіаціями рекурентних нейронних мереж, що можуть запам'ятовувати інформацію на тривалі терміни, що є важливим для розуміння змісту тексту і виявлення сентименту.

Останнім часом значну увагу привертає розвиток технологій, заснованих на механізмі уваги та трансформерах. Ці технології забезпечують моделям можливість зосередитися на найважливіших частинах тексту при їх обробці, що дозволяє досягнути значної точності у задачах, які вимагають глибокого розуміння мовних структур. Трансформери, особливо з їх двонаправленою структурою, як у BERT, здатні ефективно виконувати класифікацію тексту, розпізнавати іменовані сутності, відповідати на питання та інші задачі NLP.

Важливою складовою успіху в класифікації тексту є якість навчальних даних. Забезпечення чистих, релевантних та балансованих даних може значно покращити результати навчання моделей. Активне використання технік аугментації даних, як-от перефразування, синонімізація або зміна порядку слів у реченнях, може допомогти моделям краще загальнізувати і адаптуватися до нових даних, які вони зустрічають у реальних умовах.

Завдання класифікації тексту нерозривно пов’язане з розумінням того, як машини можуть ефективно вчитися інтерпретувати людську мову. З огляду на це, розвиток семантичного аналізу як частини процесу класифікації тексту відкриває нові шляхи для підвищення точності моделей [28]. Семантичний аналіз допомагає машинам не тільки розпізнавати слова, а й розуміти їх значення у контексті, що значно збільшує їх здатність до точної класифікації текстів за темами чи настроями.

Однією з передових технік у семантичному аналізі є розробка вбудовувань слова, які представляють слова у вигляді високорозмірних векторів, здатних захоплювати семантичні і синтаксичні взаємозв'язки між словами. Використання таких векторів дозволяє моделям глибше аналізувати текст, визначаючи залежності між словами, що з'являються у різних контекстах. Такі техніки, як Word2Vec або GloVe, забезпечують ці векторні представлення, які є основою для багатьох сучасних NLP моделей.

Також значний прогрес у класифікації тексту був досягнутий завдяки вдосконаленню алгоритмів глибокого навчання, які вміють використовувати контекстну інформацію [29]. Моделі, базовані на архітектурі Transformer, зокрема GPT та BERT, використовують контекстну інформацію для генерування тексту та його аналізу. Вони дозволяють здійснювати більш глибоке і точне розуміння тексту, розрізняючи не тільки окремі слова чи фрази, але й цілісний сенс запропонованого тексту. Застосування цих технологій виявилося революційним для багатьох областей, де потрібно автоматизувати обробку та аналіз великих обсягів тексту, таких як юридичні документи, медичні записи, наукові статті та новинні статті. Це дозволило значно

підвищити ефективність обробки інформації, роблячи можливим швидке виявлення релевантної інформації та класифікацію даних на більш високому рівні точності та релевантності.

У сфері класифікації тексту значний прогрес був досягнутий завдяки застосуванню гібридних моделей, які інтегрують різні види нейронних мереж, щоб посилити здатність моделей розуміти та класифікувати текст на глибшому рівні. Ці гібридні підходи часто поєднують конволюційні нейронні мережі (CNN), які ефективні в розпізнаванні шаблонів у тексті, з рекурентними нейронними мережами (RNN), які відомі своєю здатністю обробляти послідовні дані, зберігаючи інформацію про попередні елементи послідовності [30]. Ці гібридні моделі дозволяють використовувати переваги кожного типу мережі, забезпечуючи більш точне визначення контексту та змісту тексту. Таке поєднання може бути особливо корисним у задачах, де потрібно класифікувати текст за емоційним забарвленням або тематичною спрямованістю, забезпечуючи більш глибоке розуміння нюансів мови та більш точну реакцію на зміни в контексті. Інший важливий аспект, який варто враховувати, це застосування мета-алгоритмів у процесі класифікації тексту. Мета-алгоритми, такі як методи ансамблю, можуть значно покращити продуктивність одиночних моделей шляхом інтеграції різних стратегій навчання та предикції. Такий підхід може включати створення комбінованих прогнозів з декількох моделей для досягнення більш високої стійкості та надійності в результатах.

Автоматичне розпізнавання мови та її діалектів стає ще однією важливою областю застосування класифікації тексту, зокрема в контексті глобалізації та розширення міжнародних комунікацій [31]. Розуміння та автоматична ідентифікація різних мов та їхніх діалектів можуть значно покращити точність класифікаційних систем, особливо в застосуваннях, де важлива мультилінгвальність, таких як автоматичний переклад, мультикультурна аналітика тощо.

Ці методи та технології, у комбінації з передовими підходами до обробки та аналізу тексту, забезпечують потужні інструменти для досліджень і розвитку

у галузі машинного навчання, відкриваючи нові можливості для розуміння та використання мовних даних в різноманітних застосуваннях.

Класифікація тексту є динамічним та багатогранним напрямком у машинному навчанні, що продовжує розвиватися з впровадженням нових технологій і алгоритмів. Це забезпечує потужні засоби для обробки та аналізу текстових даних, відкриваючи нові перспективи для автоматизації та інтелектуалізації численних процесів у сучасному інформаційному суспільстві.

2.3 Метрики оцінки моделей

Використання адекватних метрик дозволяє оцінити, наскільки добре модель справляється зі своїми завданнями, та ідентифікувати можливі напрямки для покращення. Метрики оцінки в машинному навчанні забезпечують кількісне розуміння продуктивності моделей і слугують ключовим інструментом для валідації наукових досліджень та комерційних застосувань.

Однією з основних метрик для оцінки моделей класифікації є точність (Accuracy), яка вимірює частку вірно класифікованих прикладів серед усіх прикладів у тестовому наборі даних [32]. Ця метрика проста у розумінні та використанні, але може бути оманливою в незбалансованих наборах даних, де один клас значно переважає інші. Тому часто застосовують додаткові метрики, такі як точність (Precision) та відновлення (Recall). Точність показує, яка частка об'єктів, віднесених моделлю до певного класу, дійсно належить цьому класу, тоді як відновлення вимірює, яка частина об'єктів дійсно належних до класу була віднесена моделлю до цього класу. Ці метрики дозволяють глибше аналізувати продуктивність моделей, особливо в контексті задач, де важливіше виявляти певні класи, навіть за рахунок збільшення помилок у інших класах.

Комбінування точності та відновлення в одну метрику, відому як F1-оцінка, дозволяє збалансувати обидва аспекти, що є особливо корисним у випадках, коли один з показників набагато вищий за інший. F1-оцінка є гармонійним середнім між точністю та відновленням, надаючи вагу обом парамет-

рам і слугує комплексною мірою продуктивності моделі. Для задач бінарної класифікації також використовується ROC-крива (Receiver Operating Characteristic) та AUC (Area Under Curve), які аналізують взаємозв'язок між чутливістю (співвідношенням вірно позитивних результатів до суми вірно позитивних та хибно негативних результатів) і специфічністю (відношенням вірно негативних до суми вірно негативних та хибно позитивних результатів). Ці метрики важливі для оцінки здатності моделі правильно розрізняти між класами в різних порогах чутливості, що надає важливу інформацію про її поведінку в різних умовах.

Метрика як Log-Loss, або ж логарифмічна втрата, є однією з таких, яка дозволяє виміряти точність класифікаторів, надаючи більшу вагу правильним або неправильним прогнозам, залежно від їх імовірності [33]. Ця метрика особливо корисна в задачах, де необхідно мінімізувати впевнені неправильні прогнози, адже вона оцінює не просто точність категоризації, але й упевненість моделі у своїх прогнозах.

Ще одною важливою метрикою є Matthews Correlation Coefficient (MCC), яка використовується для оцінки якості бінарних класифікацій. MCC вважається однією з найбільш інформативних метрик у випадках, коли класи незбалансовані, адже вона враховує істинні та хибні позитиви та негативи в єдиній метриці, яка варіюється від -1 до +1. Метрика MCC є більш стійкою та надійною, оскільки вона враховує всі чотири квадранти конфузійної матриці, забезпечуючи більш точне зображення ефективності моделі [56]. Також, у контексті навчання з підсиленням, часто використовують метрику як Average Precision (AP), яка дозволяє оцінити класифікатор при різних рівнях відновлення. AP вимірює середню точність, отриману для кожного можливого порога класифікації, забезпечуючи більш деталізований аналіз процесу рішення, який модель застосовує при віднесенні даних до певних категорій. Це особливо корисно для розуміння, на яких порогах класифікації модель найефективніша та найточніша.

Інформаційний коефіцієнт, який також відомий як коефіцієнт взаємної інформації, вимірює кількість інформації, що ділиться між прогнозованими та

фактичними класами, та є особливо корисним у визначенні того, чи є непрямі зв'язки між змінними, що можуть впливати на продуктивність моделі.

Ще одна важлива метрика, яка часто використовується в дослідженнях, де потрібно зрозуміти варіативність відповідей моделі, – це коефіцієнт детермінації, відомий як R-квадрат. Ця метрика допомагає оцінити, яка частина змінності відгуків може бути пояснена за допомогою незалежних змінних у моделі. В контексті машинного навчання, де моделі часто стикаються з великою кількістю змінних та потенційною взаємодією між ними, R-квадрат може дати цінне уявлення про те, наскільки ефективно модель використовує доступну інформацію для прогнозування [34]. У деяких ситуаціях також може бути корисною метрика як Гейн або Lift Chart, що допомагає оцінити, наскільки модель ефективно відрізняє між різними класами порівняно з випадковим відгадуванням. Ця метрика часто застосовується в маркетингових застосуваннях для оцінки ефективності кампаній, але може бути адаптована і для класифікації тексту, де важливо визначити, наскільки вдало модель здатна ідентифікувати ключові класи серед великої кількості даних.

Часто в моделях класифікації тексту зустрічається проблема перенавчання, коли модель ідеально працює на тренувальному наборі, але її продуктивність різко падає на нових даних. Це може вказувати на надмірну адаптацію до особливостей навчального набору, які не є репрезентативними для загальної задачі.

Аналізуючи ці аспекти, можна використовувати метрики, які вимірюють величину помилок в різних частинах розподілу даних. Наприклад, розгляд величини помилок в залежності від частоти класів може виявити, чи є модель більш схильною до помилок у рідкісних класах. Це може вказувати на необхідність збалансування набору даних або внесення коректив у спосіб навчання моделі, щоб забезпечити більш однорідну продуктивність. Також значною є оцінка впливу аномалій у даних на поведінку моделі [35]. Моделі машинного навчання часто чутливі до "викидів" або аномальних значень, що можуть значно впливати на навчальний процес. Розуміння того, як модель реагує на ано-

малії, і які стратегії найкраще допомагають мінімізувати цей вплив, є важливим для підвищення її універсальності та стійкості.

Таблиця 2.2

Ключові метрики для глибшої оцінки продуктивності моделей машинного навчання

Назва метрики	Опис метрики	Контекст застосування
Log-Loss	Вимірює точність класифікаторів, надаючи вагу за точність кожного прогнозу залежно від його ймовірності.	Корисна для оцінки моделей, де важливо мінімізувати впевнені помилкові прогнози.
Matthews Correlation Coefficient (MCC)	Вимірює загальну якість бінарних класифікацій, враховуючи всі чотири квадранти конфузійної матриці.	Ідеальна для незбалансованих датасетів, оскільки вона враховує істинні та хибні позитиви і негативи.
Average Precision (AP)	Вимірює середню точність по всім рівням порогу класифікації, яка допомагає зрозуміти ефективність моделі на різних порогах чутливості.	Використовується для детального аналізу якості відгуків моделі на різних рівнях порогу.
Sensitivity to Noise	Аналізує реакцію моделі на шум у даних, ідентифікуючи стійкість до зовнішніх та внутрішніх перешкод.	Корисна для визначення стабільності моделей у реальних сценаріях використання, особливо в промислових застосуваннях.

У цій таблиці представлені метрики, які забезпечують глибший аналіз продуктивності моделей машинного навчання, з особливим акцентом на їхню чутливість до помилок та аномалій. Використання таких метрик дозволяє виявляти слабкі місця в моделях, які можуть не бути очевидними при застосуванні тра-

диційних метрик, таких як точність або відгук. Особливо важливим є розуміння того, як модель реагує на змінні умови використання та на рідкісні чи викликаючі випадки, що дозволяє адаптувати тренувальні процеси для підвищення загальної ефективності та стійкості систем.

Цей аналіз може включати розробку спеціальних метрик, які здатні ідентифікувати чутливість до конкретних видів шуму або аномалій. За допомогою цих метрик можна точніше налаштовувати моделі, розробляючи механізми, які запобігатимуть їх впливу на результати класифікації [36]. Це, в свою чергу, допомагає створювати більш робустні та надійні системи, що є ключовим для ефективного використання машинного навчання у різноманітних реальних сценаріях. Кожна з цих метрик вносить свій вклад у глибше розуміння характеристик та поведінки моделей машинного навчання, забезпечуючи необхідні інструменти для наукового аналізу та комерційного застосування. Використання цих метрик дозволяє вченим та інженерам краще розуміти та оптимізувати свої моделі, забезпечуючи більш високу точність та ефективність у вирішенні завдань класифікації та прогнозування.

Оцінка моделей за допомогою цих метрик дає можливість не тільки оцінити, наскільки добре модель виконує свої основні завдання, але й зрозуміти, як різні параметри впливають на її продуктивність. Це, в свою чергу, може слугувати основою для покращення та налагодження моделей, а також для вибору оптимальних стратегій їх застосування в реальних умовах.

Висновки до розділу 2

Розгляд основ машинного навчання забезпечує фундаментальне розуміння технік та алгоритмів, що стоять за автоматизованими системами аналізу даних. Ці знання є критично важливими для розробки та налагодження моделей, здатних ефективно обробляти та класифікувати великі обсяги текстового контенту.

Особлива увага у розділі приділена моделям для класифікації тексту, що включають засадничі підходи, такі як наївний Баєсів класифікатор, методи на основі дерев рішень та нейронні мережі. Аналіз різноманітних моделей ілюст-

рує, як кожен підхід може бути оптимізований для специфічних задач класифікації, що забезпечує не лише високу точність, але й відповідність вимогам реального часу та обмежень даних. Далі, аналіз метрик оцінки моделей надає інструментарій для критичного оцінювання і порівняння ефективності різних моделей класифікації. Вивчення таких метрик, як точність, відгук, F1-метрика, та інші, дозволяє глибше зрозуміти стійкість та надійність системи у зіткненні зі складністю реального текстового контенту. Особливо цінним є використання цих метрик для виявлення потенційних вразливостей моделей, що може привести до помилкових позитивів або негативів.

Цей розділ не тільки визначає теоретичну основу для подальших досліджень, але й підкреслює важливість ретельного вибору та налаштування моделей машинного навчання для підвищення ефективності систем виявлення фейків. Здатність до точного та швидкого аналізу тексту є ключем до розвитку надійних та ефективних інструментів, які можуть бути інтегровані у широкий спектр застосувань, від соціальних медіа до новинних платформ, де актуальність і достовірність інформації є критичною.

РОЗДІЛ 3. РОЗРОБКА ТА ВПРОВАДЖЕННЯ СИСТЕМИ ВИЯВЛЕННЯ ФЕЙКОВИХ НОВИН

3.1 Збір та підготовка даних

Збір та підготовка даних є критично важливими етапами у процесі розробки системи виявлення фейкових новин. Ефективність та точність будь-якої моделі машинного навчання безпосередньо залежать від якості та об'єму даних, на яких вона тренується. Спочатку потрібно ідентифікувати та зібрати відповідні джерела даних, які можуть включати новинні статті, блоги, повідомлення у соціальних мережах, а також коментарі та відгуки користувачів. Важливо забезпечити, щоб зібрані дані були репрезентативними та охоплювали різні перспективи та контексти, що допоможе системі ефективніше виявляти фейковий контент у різноманітних сценаріях.

Після збору ініційована підготовка даних, яка включає очищення даних від шуму та нерелевантної інформації. Це може включати видалення форматування, рекламних вставок, а також некоректних символів або посилань, які можуть ввести модель в оману. Також здійснюється нормалізація тексту, що включає перетворення всіх символів до нижнього регістру, видалення зайвих пробілів та пунктуації, а також розбивка тексту на слова або фрази, що є критично важливим для подальшої обробки. Необхідно застосувати техніки токенізації для перетворення тексту в структуру, з якою може працювати машинне навчання [37]. Це може включати використання вже готових бібліотек або розробку власних алгоритмів для специфічних потреб проекту. Важливо також виконати векторизацію тексту, перетворюючи слова в числові значення за допомогою методів, таких як bag-of-words або TF-IDF, що дозволяє машинній моделі ефективно працювати з текстовими даними.

Ще одним кроком у підготовці даних є розділення набору даних на тренувальну, тестову та валідаційну вибірки. Це дозволяє не тільки адекватно навчити модель, але й перевірити її здатність до узагальнення на нових, раніше не

бачених даних. Ефективне розділення даних допомагає уникнути перенавчання моделі та забезпечує її надійність та об'єктивність у реальних умовах.

Під час розробки та впровадження системи для виявлення фейкових новин, збір та підготовка даних відіграють ключову роль у забезпеченні високої якості та ефективності майбутньої моделі. Крім загальних технік очищення та нормалізації тексту, особлива увага приділяється анотації даних та визначенню зразків даних, які можуть бути використані для тренування та перевірки моделі. Для цього важливо розробити чіткі критерії, за якими новини класифікуються як 'правдиві' або 'фейкові', а також забезпечити, щоб датасет був збалансованим відносно цих категорій.

Ефективна підготовка даних також вимагає залучення експертів із засобів масової інформації та фактчекінгу, які можуть допомогти валідувати та верифікувати дані, забезпечуючи їхню вірність та релевантність. Використання автоматизованих інструментів для попереднього вибору та аналізу змісту новин може суттєво підсилити процес підготовки даних, дозволяючи швидше ідентифікувати потенційні фейки та забезпечити об'єктивність класифікації. Для більш глибокого аналізу тексту, техніки обробки природної мови, такі як розпізнавання іменованих сутностей, аналіз тональності та семантичний аналіз, можуть бути впроваджені для виявлення тонких нюансів та контекстуальних ознак у текстах, які часто вказують на фейковий контент [38]. Застосування машинного навчання до цих складних задач дозволяє моделям вчитися розпізнавати складні патерни та аномалії в текстах, що значно підвищує точність виявлення фейкових новин.

Інтеграція різноманітних джерел даних, включаючи історичні дані, дані соціальних медіа та інші форми цифрового контенту, може допомогти системі виявлення фейків адаптуватися до змінюваних методів та стратегій, які використовуються для створення та поширення неправдивої інформації. Це, в свою чергу, забезпечує більш глибокий контекстуальний аналіз та допомагає системі виявляти нові та еволюціонуючі форми фейкових новин.

В рамках збору та підготовки даних для системи виявлення фейкових новин, особливу увагу слід звернути на створення та використання авансованих алгоритмів для автоматичного розпізнавання складних мовних зворотів, які часто використовуються в маніпулятивних цілях. Використання технологій глибокого навчання та нейронних мереж дозволяє ефективно обробляти великі обсяги текстових даних та виділяти приховані взаємозв'язки та патерни, які можуть не бути очевидними для традиційних методів аналізу.

Важливо також враховувати культурний та географічний контекст, у якому генерується інформація, адже фейкові новини часто мають регіональні особливості, що можуть впливати на манеру подачі та зміст інформації. Інтеграція таких аспектів у систему допомагає підвищити її чутливість та специфічність при ідентифікації неправдивих повідомлень [39]. Процес збору даних не має обмежуватись лише збіром текстових новин. Включення візуального контенту, аудіо записів та відеоматеріалів може значно розширити потенціал системи у виявленні фейкових новин. Аналіз зображень та відео за допомогою комп'ютерного зору та технологій розпізнавання облич може виявити дезінформацію, яка може бути непомітною виключно у текстовому форматі.

Подальше вдосконалення системи передбачає використання адаптивних моделей, здатних самонавчання та адаптації до постійно змінюваних умов інформаційного простору. Це передбачає створення механізмів швидкого оновлення даних і алгоритмів, а також залучення користувачів до процесу верифікації інформації, що в свою чергу може допомогти підвищити загальну надійність системи.

В рамках розробки системи для виявлення фейкових новин особливо важливим є впровадження методів машинного навчання, що дозволяють автоматизувати процеси аналізу емоційного забарвлення тексту. Використання аналізу сентименту може значно підвищити ефективність систем, здатних ідентифікувати маніпулятивні повідомлення, які часто використовують емоційно заряджені висловлювання для впливу на аудиторію [40]. Автоматизований аналіз тексту, заснований на визначенні ключових емоційних слів та фраз, дозволяє

швидко оцінювати тон повідомлень і виділяти потенційно маніпулятивний контент.

Для забезпечення більш глибокого розуміння контексту повідомлень, система може включати компоненти, які аналізують зв'язки між різними новинними джерелами та їх історичну достовірність. Це включає в себе вивчення звичайних патернів поширення інформації серед певних джерел, виявлення аномалій у способах подачі новин, які можуть вказувати на спроби маніпуляції громадською думкою. Інтеграція технологій обробки природної мови (NLP) заснована на алгоритмах, які можуть розпізнавати складні мовні конструкції та відтворювати семантичні структури тексту. Це дозволяє системі розпізнавати складні інформаційні кампанії, що використовують субтільні лінгвістичні техніки для впливу на сприйняття фактів.

Використання розширених баз даних, які містять перевірені факти, та алгоритми порівняння фактичної інформації з цими базами дозволяє системі швидко ідентифікувати недостовірні дані, що є ключовим елементом в боротьбі з дезінформацією. Важливо також створити механізми оновлення та розширення баз даних, щоб система могла адаптуватися до постійно змінюваних методів поширення фейкових новин.

Таблиця 3.1

Компоненти та функції системи виявлення фейкових новин

Компонент	Функції	Роль в системі
Аналіз сентименту	Визначення емоційного забарвлення тексту, виявлення маніпулятивних повідомлень.	Допомагає виявляти текст з надмірно емоційним забарвленням, який часто використовується у фейкових новинах.
Історична верифікація	Порівняння заяв з історично достовірними даними, використання баз	Забезпечує перевірку фактів та допомагає ідентифікувати недостовірну інформацію.

	даних перевірених фактів.	
Контекстуальний аналіз	Розширений аналіз тексту для виявлення складних мовних патернів та зв'язків.	Підвищує точність системи за рахунок виявлення прихованих смислів і маніпуляцій.
Обробка природної мови	Розпізнавання мовних конструкцій, семантичний аналіз, визначення структур мовлення.	Поліпшує розуміння контенту, забезпечує глибший аналіз тексту для виявлення фейкового контенту.
Інтеграція різноманітних джерел	Збір інформації з множини платформ та форматів, включно з текстовими, візуальними та аудіо даними.	Дозволяє комплексний аналіз різних типів контенту для всебічного виявлення фейків.

Таблиця 3.1 демонструє, як різноманітні компоненти системи взаємодіють для ефективного виявлення фейкових новин. Кожен компонент виконує специфічні функції, які сприяють загальній точності та надійності системи. Наприклад, інтеграція аналізу сентименту може допомогти визначити потенційно маніпулятивний контент, в той час як історична верифікація та контекстуальний аналіз гарантують, що інформація перевірена та достовірна. Разом, ці компоненти формують комплексну систему, здатну адаптуватися до швидкозмінних викликів сучасної інформаційної ери та ефективно боротися з фейковими новинами.

Розробка ефективної системи виявлення фейкових новин вимагає комплексного підходу, який поєднує передові технології обробки даних з глибоким розумінням соціальних та культурних аспектів, що впливають на сприйняття інформації.

Етап збору та підготовки даних вимагає ретельного планування та виконання. Забезпечення якості даних, їхньої валідності та репрезентативності є критично важливим для створення ефективної системи виявлення фейкових но-

вин [41]. Ці кроки визначають основу для створення надійної та точної моделі, яка може ефективно боротися з викликами сучасної інформаційної ери. Такий підхід до збору та підготовки даних є фундаментальним для створення надійної системи виявлення фейкових новин, яка може ефективно функціонувати у широкому діапазоні контекстів і викликів сучасного інформаційного простору.

3.2 Навчання та налаштування моделі

Процес навчання та налаштування моделі в системі виявлення фейкових новин є одним із найважливіших етапів розробки, оскільки саме від якості та точності навчання залежить ефективність всієї системи. Цей процес починається з вибору відповідної моделі машинного навчання, яка може бути базована на методах класичної статистики, як-то наївний Баєс або логістична регресія, або на більш складних алгоритмах, таких як нейронні мережі або машини опорних векторів.

Основною задачею на етапі навчання є визначення параметрів моделі, які найкраще описують закономірності у даних, що дозволять розрізняти правдиві новини від фейкових з високою точністю. Для цього необхідно забезпечити, що навчальний датасет містить достатньо прикладів обох класів, а також що він є репрезентативним, тобто охоплює широкий спектр можливих сценаріїв, з якими модель може зіткнутися в реальному світі [55]. На наступному кроці важливо провести тонке налаштування параметрів моделі, включаючи складність моделі, швидкість навчання, і кількість ітерацій навчання. Це може вимагати численних експериментів і тестувань з використанням крос-валідації або валідаційних наборів даних, щоб знайти оптимальний баланс між здатністю моделі генералізувати дані та її точністю на конкретних прикладах.

Одним із критичних аспектів навчання є запобігання перенавчанню, коли модель занадто добре адаптується до даних навчання та втрачає здатність генералізації на нових даних [42]. Для цього можуть бути використані методи, такі як регуляризація, dropout для нейронних мереж, або введення штрафних

функцій для складності моделі. Такі методи допомагають зменшити ризик перенавчання та підвищити здатність моделі до праці з реальними, часто непередбачуваними даними. Ще один важливий етап проводиться оцінка моделі на тестовому наборі даних, який не використовувався під час навчання. Це дозволяє переконатися в тому, що модель має достатню точність і готова до впровадження в реальні умови. Оцінка моделі включає аналіз її продуктивності з використанням метрик, таких як точність, відгук, F1-метрика, а також розгляд її поведінки на прикладах різних класів.

Процес навчання та оптимізації моделі в системі виявлення фейкових новин може стати значно ефективнішим завдяки застосуванню передових методів збору даних та їх аналізу. Один із ключових аспектів, який потребує уваги, це розробка механізмів для автоматичного оновлення та поповнення даних, які використовуються для тренування моделі. Це включає інтеграцію з онлайн-джерелами, які неперервно надають свіжі дані, забезпечуючи тим самим, що система залишається актуальною і ефективною перед обличчям швидкозмінних інформаційних кампаній та нових методів маніпуляції.

Застосування методів штучного інтелекту, таких як техніки глибокого навчання з використанням автоенкодерів або генеративно-змагальних мереж, може значно підвищити здатність системи ідентифікувати складні та тонкі патерни у даних, які вказують на фейковий контент. Ці технології дозволяють моделі вчитися на більш глибокому рівні, розпізнавати неочевидні зв'язки та аномалії, які можуть виявити приховані спроби маніпуляції. Інтеграція мультидатачних аналітичних платформ дозволяє об'єднувати текстові дані з відео, аудіо та зображеннями, надаючи системі комплексний набір інструментів для аналізу. Це зокрема важливо у контексті розпізнавання фейків, оскільки дезінформація часто включає в себе складні мультимедійні елементи, які можуть бути використані для підсилення впливу текстового повідомлення.

Для оптимізації процесу навчання моделі необхідно також впровадити системи раннього попередження, які здатні ідентифікувати потенційні проблеми у навчанні, такі як перенавчання або недостатнє навчання [43]. Ці системи

моніторингу допоможуть вчасно коригувати параметри навчання, а також забезпечувати баланс між здатністю моделі адаптуватися до нових даних та зберігати загальну стабільність та високу продуктивність.

Під час навчання та налаштування моделі в рамках системи виявлення фейкових новин, критично важливою є фаза валідації моделі, яка вимагає глибокого розуміння як даних, так і потенційних зовнішніх впливів. Валідація моделі забезпечує не лише підтвердження її здатності точно класифікувати дані, але й визначення стійкості моделі до змін у вхідних даних, що може бути викликано еволюцією методів створення фейкових новин.

Інтеграція сценаріїв стрес-тестування моделі також є важливою частиною процесу налаштування. Ці сценарії дозволяють визначити, як модель реагує на несподівані або крайні умови, наприклад, коли вона стикається з даними, які значно відрізняються від тих, що були представлені під час навчання. Це включає в себе введення великої кількості шуму або систематично змінених даних для перевірки гнучкості і адаптивності моделі.

Для ефективного налаштування моделі, критично важливим є застосування технік оптимізації гіперпараметрів, таких як кількість шарів у нейронній мережі, її архітектура, швидкість навчання та інші. Оптимізація гіперпараметрів може бути виконана за допомогою методів, таких як сітковий пошук, випадковий пошук або більш складних алгоритмів, таких як байєсівська оптимізація. Ці методи дозволяють систематично оцінити велику кількість комбінацій налаштувань і вибрати ті, що найкраще працюють для конкретної задачі.

В процесі розробки системи виявлення фейкових новин, одним із критичних етапів є адаптація моделі до змінюваних умов інформаційного середовища [44]. Це включає не лише первинне навчання моделі, але й її постійне оновлення та модернізація для відповідності актуальним викликам. Використання методів трансферного навчання може значно підвищити адаптивність моделі, дозволяючи їй використовувати здобуті знання в одному контексті для швидкого та ефективного вирішення задач в інших, нових контекстах. Цей підхід може допомогти моделі краще реагувати на еволюцію мовних шаблонів та стратегій

поширення дезінформації. Ключовим аспектом налаштування моделі є також впровадження механізмів виявлення аномалій, що дозволяють системі ідентифікувати і аналізувати відхилення від звичайних моделей поведінки даних. Ці технології здатні виявляти не лише явні спроби маніпуляції, але й тонкі маніпулятивні впливи, які можуть виявлятися у вигляді незначних модифікацій у використанні мови або презентації інформації.

Залучення і використання великих даних для тренування та налаштування моделі має вирішальне значення, оскільки дозволяє моделі вчитися на різноманітних і комплексних датасетах, що відображають реальне інформаційне середовище. Інтеграція даних з різних джерел та мовних груп дозволяє створювати більш універсальні моделі, спроможні ефективно діяти в глобальному інформаційному просторі. Важливою є взаємодія моделі з користувачами, які можуть надавати зворотний зв'язок про ефективність її роботи. Це не лише підвищує довіру до системи, але й сприяє її постійному вдосконаленню через корекцію помилок та недоліків, виявлених користувачами. Такий підхід допомагає системі стати більш гнучкою і адаптованою до потреб користувачів.

Для забезпечення високої ефективності системи виявлення фейкових новин, навчання та налаштування моделі має включати розробку імплементації когнітивного аналізу, який бере до уваги не лише мовні показники, але й контекст, в якому інформація була створена і розповсюджена [45]. Цей підхід дозволяє системі краще розуміти нюанси інтенцій за текстами, використовуючи передові методи аналізу поведінкових патернів, які могли б вказувати на недобросовісність джерела.

Значну роль у навчанні моделі відіграє також впровадження технологій великих даних, які дозволяють обробляти широкий спектр інформації з різних джерел в реальному часі. Використання біг дата технологій забезпечує не тільки здатність до швидкої обробки даних, але й здатність знаходити зв'язки між різноманітними даними, які можуть не бути очевидними при більш традиційному аналізі. Застосування інноваційних методів машинного навчання, таких як згорткові нейронні мережі (CNN) для аналізу текстів, та рекурентні нейронні

мережі (RNN) для виявлення та аналізу секвенційних даних, може значно покращити здатність системи ідентифікувати тонкі маніпуляції в тексті. Ці методи навчання забезпечують глибший рівень аналізу і більш точне розуміння структури та змісту інформації.

Практика реалізації постійного оновлення баз даних з прикладами фейкових та правдивих новин є критичною для адаптації моделі до змінюваних методів маніпуляції. Створення адаптивної системи, що здатна вчитися з нових даних та оновлювати свої алгоритми відповідно до актуальних трендів, забезпечує тривалу ефективність і релевантність в реальному світі.

Включення механізмів зворотного зв'язку від користувачів системи є життєво важливим для остаточного налаштування та покращення моделі. Користувачі можуть надати цінну інформацію про ефективність моделі в реальних умовах, вказуючи на типи помилок, частоту виникнення фейкових позитивних або негативних результатів, а також можливі недоліки в здатності моделі адаптуватися до нових видів дезінформації. Цей зворотний зв'язок може бути використаний для подальшої фінтюнінгу моделі, забезпечуючи її надійність і точність в довгостроковій перспективі.

Процес навчання та оптимізації моделі в системі виявлення фейкових новин вимагає не тільки використання передових технологій машинного навчання, але й постійне оновлення підходів і методологій з урахуванням нових викликів і загроз в галузі інформаційної безпеки [46]. Цільова розробка і тонке налаштування моделі виявлення фейкових новин є комплексним процесом, що вимагає глибокого аналізу та оптимізації, з метою забезпечення високої надійності та ефективності в реальних операційних умовах.

3.3 Впровадження та тестування системи

Впровадження та тестування системи виявлення фейкових новин є завершальним та одним з найважливіших етапів у процесі розробки. Після розробки та налаштування моделі наступним кроком є її інтеграція у реальне середови-

ще, де вона буде функціонувати. Цей процес включає кілька ключових компонентів, які забезпечують плавний перехід від теорії до практики, зокрема, деплоймент моделі, моніторинг її роботи та періодичні оцінки її ефективності.

Перед запуском системи у виробництво, вона проходить ретельне тестування, що включає кілька етапів. Початково проводиться тестування у контрольованому середовищі, щоб виявити будь-які технічні помилки або вразливості. Це включає в себе юніт-тести окремих компонентів, інтеграційне тестування їх взаємодії та системне тестування повної архітектури системи. Важливою частиною тестування є також перевірка відповідності системи вимогам безпеки та приватності, що особливо критично для систем, що обробляють чутливу інформацію.

Після внутрішніх тестів система може бути випробувана у пілотному проекті з реальними користувачами на обмеженій території або серед обмеженої аудиторії. Пілотні тести дозволяють оцінити, як система працює в реальних умовах, та зібрати зворотний зв'язок від реальних користувачів. Це критично для забезпечення того, що система є корисною, зручною в експлуатації та ефективною відповідно до свого призначення [47]. Важливою частиною процесу впровадження є налагодження моніторингу системи, що включає в себе слідкування за її продуктивністю, часом відгуку та будь-якими непередбаченими помилками, що можуть виникнути. Системи моніторингу та сповіщення дозволяють оперативно реагувати на проблеми, забезпечуючи високу доступність та надійність системи. Залучення аналітики даних та використання засобів візуалізації допомагають зрозуміти більш глибокі тренди у функціонуванні системи та ефективно адаптувати її до змінних умов експлуатації. Важливим етапом є проведення оцінки ефективності системи на основі зібраних під час експлуатації даних. Це включає аналіз помилок, оцінку точності виявлення фейків, та інші метрики, які можуть бути визначені на етапі проектування. Результати цієї оцінки використовуються для подальшого удосконалення системи, що може включати налаштування параметрів моделі, оновлення алгоритмів або навіть зміну архітектури системи в цілому.

У процесі впровадження та тестування системи виявлення фейкових новин, існує важливість створення стратегій для безперервного оновлення та адаптації моделі до нових викликів інформаційного простору. Основним завданням є забезпечення того, щоб система залишалася релевантною та ефективною навіть у випадку зміни тактик поширення дезінформації або введення нових форм медійного контенту, що вимагає постійного моніторингу тенденцій та інновацій у сфері цифрових медіа.

Одним з ключових елементів цього процесу є використання технологій штучного інтелекту для аналізу поведінкових шаблонів користувачів та автоматичного виявлення потенційно шкідливого контенту [54]. Це передбачає застосування машинного навчання для аналізу великих масивів даних, включаючи тексти новин, соціальні мережі, блоги, та інші джерела, де фейкові новини можуть бути поширені. Застосування глибокого аналізу даних допомагає ідентифікувати не лише явні випадки дезінформації, але й більш тонкі спроби маніпуляцій громадською думкою.

Ключовою стратегією для успішного тестування та впровадження є створення симуляційних середовищ, де можна випробувати систему у контрольованих, але реалістичних умовах. Це дозволяє не тільки перевірити технічну спроможність системи, але й оцінити її ефективність у різних сценаріях використання. Симуляції можуть включати стрес-тести, які імітують високу інтенсивність інформаційних атак, або тести стійкості, що дозволяють визначити здатність системи працювати надійно під час технічних збоїв або при великому навантаженні.

Роль зворотного зв'язку від користувачів є невід'ємною частиною процесу тестування та впровадження. Збір відгуків допомагає виявити недоліки системи, які можуть не бути очевидні під час внутрішніх тестів. Це також включає аналіз поведінки реальних користувачів, що може допомогти оптимізувати інтерфейс та функціонал системи для забезпечення її зручності та ефективності в реальних умовах.

Ефективне впровадження та тестування системи виявлення фейкових новин вимагає забезпечення її гнучкості та масштабованості, щоб вона могла адаптуватися до непередбачуваних змін у ландшафті інформації. Це означає, що система має бути здатна швидко обробляти великі обсяги даних з різноманітних джерел і в різних форматах, включаючи текст, зображення, відео та аудіо [48]. Розробка такої системи включає використання передових обчислювальних технологій та алгоритмів, які можуть динамічно масштабуватися залежно від потреб.

Забезпечення міцної інфраструктури, яка підтримує безперебійну роботу системи, є життєво важливим. Використання хмарних технологій дозволяє не тільки гарантувати високу доступність та відмовостійкість системи, але й забезпечити її еластичність для обробки сплесків даних під час великих новинних подій або інформаційних атак. Хмарні платформи також надають можливості для використання передових інструментів моніторингу та аналітики, що допомагають відстежувати продуктивність системи та швидко реагувати на потенційні проблеми.

Важливим аспектом тестування системи є розробка детальних сценаріїв тестування, які імітують реальні оперативні умови. Ці сценарії повинні включати не тільки стандартні випадки використання, але й крайові випадки та потенційні атаки, що спрямовані на систему. Використання технік штучного інтелекту для генерації синтетичних даних може допомогти створити різноманітні тестові сценарії, які допоможуть краще оцінити стійкість системи до спроб маніпуляції. Крім технічного тестування, необхідно проводити оцінку впливу системи на суспільство та її прийняття користувачами. Це включає збір зворотного зв'язку від реальних користувачів, аналіз їх вражень від взаємодії з системою та вивчення їхніх занепокоєнь щодо приватності, точності та надійності [49]. Залучення зацікавлених сторін, таких як журналісти, правозахисники та освітяни, може допомогти зрозуміти більш широкий соціальний контекст, в якому система буде використовуватися, та оптимізувати її для досягнення балансу між ефективністю та етичністю в її застосуванні.

Система має бути здатною інтегруватися з існуючими медіа-платформами, соціальними мережами, та іншими цифровими сервісами, забезпечуючи безперервний аналіз та моніторинг інформації. Така інтеграція вимагає розробки API, які дозволять забезпечити легку взаємодію системи з іншими програмними продуктами, що значно розширює її функціональні можливості та сферу застосування.

Під час тестування системи особливу увагу потрібно приділити перевірці її здатності виявляти фейкові новини в реальному часі. Це включає оцінку швидкості обробки даних та часу реакції на нові інформаційні потоки, що є критично важливим для оперативності відгуків системи на постійно змінювані новинні події. Іншим важливим аспектом є забезпечення стабільності системи під час пікових навантажень, що може включати адаптацію обчислювальної інфраструктури під час великих інформаційних кампаній або глобальних подій. Також критично важливою є здатність системи адаптуватися до нових видів фейків, які постійно еволюціонують [50]. В цьому контексті, система повинна мати вбудовані механізми машинного навчання, які дозволяють їй постійно навчатися з отриманих даних та вдосконалювати свої алгоритми для ідентифікації нових шаблонів дезінформації. Регулярне оновлення бази даних з прикладами фейкових новин, включаючи збір та аналіз останніх випадків, допоможе системі залишатися актуальною в умовах, коли методи маніпуляції інформацією постійно удосконалюються. Важливою складовою процесу тестування є також проведення етичної оцінки системи, зокрема аналіз потенційного впливу на права та свободи користувачів. Це включає переконання в тому, що система не порушує конфіденційність даних та не веде до необґрунтованої цензури. Розробка чітких критеріїв для оцінки інформації та прозорість процесів прийняття рішень системою є важливими для забезпечення її суспільної прийнятності та довіри.

При впровадженні та тестуванні системи виявлення фейкових новин важливо забезпечити, що процес є ітеративним та включає постійний зворотний зв'язок від користувачів та аналітиків, що спостерігають за системою. Це дозво-

ляє системі вчасно адаптуватися до нових викликів та трендів дезінформації, що є особливо важливим у світі, де методи маніпуляції інформацією швидко еволюціонують. Під час цього процесу особливу увагу слід приділяти не лише збору та аналізу кількісних даних про ефективність системи, але й якісним оцінкам її впливу на користувачів та інформаційне середовище загалом.

Ще одним критичним аспектом є розробка механізмів для забезпечення прозорості та зрозумілості рішень, прийнятих системою. Це включає створення інтерфейсів, що дозволяють користувачам бачити, чому певна інформація була класифікована як фейкова, з можливістю перегляду доказів або пояснень, які підкріплюють це рішення [51]. Важливість цього аспекту не можна недооцінювати, оскільки він сприяє довірі до системи та її прийняттю користувачами. Крім технічного тестування, система повинна пройти юридичну та етичну перевірку, щоб забезпечити відповідність всім вимогам та нормам. Це включає аналіз можливих правових наслідків застосування такої системи, включаючи питання цензури, втручання в приватне життя та можливість помилкової класифікації інформації. Передбачення та мінімізація таких ризиків через дизайн системи та її політику використання є ключовими для створення відповідальної та соціально корисної технології.

На завершальному етапі впровадження системи важливо також забезпечити її інтеграцію з існуючими інформаційними системами та платформами. Це дозволяє системі ефективно функціонувати в реальних умовах, забезпечуючи швидке реагування на появу фейкових новин та їх нейтралізацію. Залучення партнерів із медіа індустрії, технологічних компаній та наукових установ може сприяти ширшому впровадженню системи, збільшуючи її вплив та ефективність у боротьбі з дезінформацією.

Особлива увага приділяється процесу її аудиту та верифікації за допомогою зовнішніх незалежних організацій. Це дозволяє не лише підтвердити відповідність системи всім встановленим стандартам, але й забезпечує додатковий рівень довіри до її результатів з боку користувачів. Незалежний аудит

включає ретельний перегляд алгоритмів, використаних у системі, аналіз даних, на яких вона була навчена, і методів, які використовуються для її оновлення.

Ефективне впровадження системи потребує розробки детального плану на випадок потенційних проблем, які можуть виникнути під час її експлуатації [53]. Це включає створення процедур швидкого реагування на помилкові спрацьовування системи, розробку механізмів виправлення помилок, та стратегій комунікації з користувачами щодо причин та методів вирішення проблем. Подібні заходи допомагають підтримувати високу репутацію системи та її ефективність на довготривалу перспективу. Для забезпечення неперервної ефективності системи, важливим є також впровадження постійного процесу її оновлення та вдосконалення. Це означає систематичний аналіз зібраних даних про її роботу, вивчення нових методів дезінформації, які з'являються у світі, та оновлення алгоритмів відповідно до останніх наукових досліджень у галузі штучного інтелекту та машинного навчання. Впровадження адаптивного навчання, де система самостійно вчиться на нових даних та виправляє свої помилки, може значно підвищити її здатність швидко реагувати на зміни у патернах дезінформації.

Таблиця 3.2

Ключові етапи впровадження та тестування системи виявлення фейкових новин

Етап	Опис діяльності
Незалежний аудит	Перевірка алгоритмів та даних системи незалежними експертами для забезпечення об'єктивності та

	відповідності стандартам.
Розробка стратегій реагування на проблеми	Створення процедур для швидкого вирішення помилок системи та забезпечення неперервності її роботи.
Адаптація та оновлення алгоритмів	Постійне оновлення алгоритмів системи для відповідності новим методам маніпуляції інформацією.
Юридична та етична перевірка	Аналіз потенційних правових та етичних ризиків, пов'язаних з використанням системи.
Інтеграція з існуючими системами	Забезпечення сумісності системи з іншими інформаційними платформами для гармонійної інтеграції.
Проведення інформаційної кампанії	Організація освітніх та промоційних заходів для залучення користувачів та підвищення обізнаності про систему.

Таблиця 3.2 систематизує ключові етапи, які необхідно врахувати при впровадженні та тестуванні системи виявлення фейкових новин. Вона демонструє взаємозв'язок між технічними, юридичними, етичними та комунікаційними аспектами проекту, забезпечуючи цілісний підхід до реалізації системи. Кожен етап має своє значення для гарантування ефективності, безпеки, та прийнятності системи серед кінцевих користувачів. Це важливо для забезпечення не тільки технічної спроможності системи ефективно ідентифікувати фейкові новини, але й для створення відкритої та відповідальної інформаційної среди.

Для забезпечення широкого впровадження та використання системи, необхідно провести обширну інформаційну кампанію серед потенційних користувачів, яка б включала освітні програми, демонстрації можливостей системи та її переваг [52]. Така кампанія повинна включати не лише технічні аспекти, але й висвітлювати соціальні аспекти використання системи, такі як боротьба з

дезінформацією, підвищення інформаційної грамотності населення, та зміцнення демократичних процесів через підтримку достовірної інформації.

Такий комплексний підхід до впровадження та тестування забезпечує створення ефективної, безпечної та прийнятної системи виявлення фейкових новин, що може служити важливим інструментом у забезпеченні інформаційної достовірності та підтриманні демократичних процесів. Процес впровадження та тестування системи виявлення фейкових новин є складним і багатогранним завданням, яке вимагає інтегрованого підходу та постійного вдосконалення для забезпечення її адекватності та ефективності у швидко змінюваному інформаційному просторі.

Висновки до розділу 3

Було ретельно розглянуто критичні аспекти від збору і підготовки даних, через навчання та налаштування моделі, до фінальних етапів її впровадження та тестування. Цей процес вимагає не тільки технічної експертизи, але й глибокого розуміння соціальних та етичних наслідків застосування таких систем.

Збір та підготовка даних виявляються критичними для забезпечення якості та релевантності інформації, яка використовується для тренування моделі. Процес охоплює збір відповідних, диверсифікованих та об'єктивних даних, які мають стати основою для виявлення фейкових новин. Підготовка даних включає їх очищення, нормалізацію та анотацію, що допомагає підвищити точність майбутньої моделі. Етап навчання та налаштування моделі вимагає застосування складних алгоритмів машинного навчання та штучного інтелекту, а також тонкого налаштування гіперпараметрів, щоб забезпечити оптимальну роботу системи. Важливість цього етапу полягає у здатності моделі адаптуватися та ефективно реагувати на постійно змінювані методи маніпуляції інформацією. Впровадження та тестування системи, забезпечує перевірку її ефективності у реальних умовах та оцінку здатності системи до ідентифікації фейкових новин. Впровадження системи включає її інтеграцію з різними інформаційними плат-

формами та медійними ресурсами, а також постійний моніторинг її продуктивності для своєчасного виявлення та вирішення будь-яких проблем. Тестування охоплює верифікацію функціональності, безпеки, та надійності системи, а також включає аналіз її прийнятності користувачами та впливу на суспільство.

Разом ці етапи формують стратегічний та системний підхід до створення та впровадження систем виявлення фейкових новин, що не тільки підвищує інформаційну грамотність і захист споживачів інформації, але й сприяє створенню більш прозорого та відповідального інформаційного середовища.

ВИСНОВКИ

Розділ 1 глибоко занурюється у феномен фейкових новин, досліджуючи їх концептуалізацію та типологію. Вивчення літератури з цієї проблематики виявило складність феномена, вказуючи на різноманіття форм і методів, якими фейкові новини можуть впливати на громадську думку. Обговорення методів виявлення фейкових новин в цьому розділі виокремило ключові підходи та технології, що застосовуються в сучасній інформаційній аналітиці, включаючи як мануальні, так і автоматизовані методи. Також розділ наголошує на основних викликах та проблемах у боротьбі з дезінформацією, зокрема на складнощах визначення вірогідності джерел і верифікації фактів у швидкоплинному інформаційному потоці.

Розділ 2 концентрується на теоретичних основах машинного навчання, які є критично важливими для розробки ефективних алгоритмів класифікації та аналізу тексту. Вивчення основ машинного навчання дало змогу зрозуміти, як різні моделі можуть бути налаштовані для виявлення складних мовних патернів, які часто характеризують маніпулятивні повідомлення. Далі, розгляд моделей для класифікації тексту виявив потенціал застосування різних алгоритмів, від простих до більш складних нейронних мереж, для обробки і аналізу текстових даних на предмет їхньої достовірності. Окрім того, детальний аналіз метрик оцінки моделей підкреслив значення точності, повноти та інших критеріїв, які дозволяють оцінити якість та ефективність використаних алгоритмів.

Ці два розділи разом закладають міцну основу для розуміння складності проблеми фейкових новин та надають інструменти та методологію для їх виявлення і аналізу, що є критично важливим для подальшої розробки та впровадження системи детекції.

Розділ 3 являє собою кульмінаційний етап, де зібрані теоретичні знання та методологічні підходи, обговорені в попередніх розділах, інтегруються і застосовуються для розробки та впровадження реальної системи виявлення фейкових новин. Цей розділ детально описує весь процес від збору та підготовки да-

них, через навчання та налаштування моделі, до її впровадження та тестування в реальних умовах.

Збір та підготовка даних вимагають ретельної організації та виконання, оскільки якість вхідних даних безпосередньо впливає на результативність системи. Цей етап включає ідентифікацію та збір релевантних датасетів, їх очищення від шуму та відповідну анотацію. Важливим є також забезпечення репрезентативності даних, щоб система могла ефективно працювати з різними типами контенту. Навчання та налаштування моделі вимагають глибокого розуміння властивостей використовуваних алгоритмів та їх адаптації до специфічних завдань і даних. На цьому етапі розробники використовують методи машинного навчання для того, щоб "навчити" систему розрізняти достовірні новини від фейкових. Важливим аспектом є оптимізація моделі для досягнення найкращого балансу між точністю виявлення та швидкістю обробки, що важливо для реальної експлуатації системи.

Впровадження та тестування системи — це фінальні кроки, які забезпечують перехід від прототипу до робочої системи. Впровадження передбачає інтеграцію системи у реальне середовище, де вона буде використовуватися. Важливою складовою цього процесу є тестування системи на стабільність та надійність у різних умовах, а також перевірка її здатності адекватно реагувати на непередбачені ситуації та атаки дезінформації. Крім того, оцінка зворотного зв'язку від користувачів допомагає вдосконалити систему, підвищуючи її ефективність та користувацьку прийнятність.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Аксенов І. В., Сидорова Л. М. Використання нейронних мереж для класифікації текстових даних (дата звернення:03.06.2024).

2. Андрієнко І. О., Петрова А. В. Аналіз методів машинного навчання для виявлення фейкових новин (дата звернення:01.06.2024).

3. Білецький В. С., Григор'єва Ю. П. Моделювання процесів виявлення фейкової інформації (дата звернення:01.06.2024).

4. Богданович О. М., Кузьмін М. В. Методи текстової класифікації в контексті інформаційної безпеки (дата звернення:01.06.2024).

5. Воробйов Д. П., Журавльов О. К. Системи автоматичного виявлення дезінформації (дата звернення:01.06.2024).

6. Гаврилюк М. С., Іванова Л. П. Огляд сучасних підходів до виявлення фейкових новин (дата звернення:01.06.2024).

7. Гончаренко В. В., Шевченко І. М. Використання алгоритмів машинного навчання для аналізу тексту (дата звернення:01.06.2024).

8. Гриценко А. В., Тимофєєв В. М. Механізми детекції фейкових новин у соціальних мережах (дата звернення:01.06.2024).

9. Даниленко О. П., Кравець Ю. С. Інтелектуальні системи аналізу текстової інформації (дата звернення:04.06.2024).

10. Дмитренко Н. І., Король А. В. Використання машинного навчання для автоматизації процесів виявлення фейкових новин (дата звернення:01.06.2024).

11. Дудник О. Л., Павленко Т. Г. Методи класифікації тексту з використанням машинного навчання (дата звернення:01.06.2024).

12. Єфремов А. В., Литвиненко М. С. Моделі та методи виявлення дезінформації у текстовому контенті (дата звернення:01.06.2024).

13. Жуковська І. О., Коваленко П. В. Аналіз методів обробки текстових даних для виявлення фейків (дата звернення:01.06.2024).

14. Задорожний М. В., Ткаченко І. І. Інформаційна безпека та боротьба з фейковими новинами (дата звернення:02.06.2024).

15. Іващенко П. М., Чумаченко А. В. Використання нейромереж для виявлення дезінформації у текстах (дата звернення:01.06.2024).

16. Ільченко О. Г., Кириленко В. М. Методи машинного навчання для класифікації тексту (дата звернення:01.06.2024).

17. Коваленко В. І., Мельник Л. П. Аналіз ефективності алгоритмів машинного навчання для виявлення фейкових новин (дата звернення:01.06.2024).

18. Ковтун В. О., Сидоренко Т. М. Системи автоматичного виявлення фейкових новин (дата звернення:03 .06.2024).

19. Козлов В. С., Пархоменко М. В. Використання методів глибокого навчання для аналізу текстових даних (дата звернення:01.06.2024).

20. Коротун І. М., Мазур І. В. Методи текстової аналітики для виявлення дезінформації (дата звернення:04.06.2024).

21. Кравченко О. П., Шевчук М. Г. Використання методів машинного навчання для класифікації текстових даних (дата звернення:01.06.2024).

22. Краснов В. І., Тарасов В. В. Методи виявлення фейкових новин у соціальних медіа (дата звернення:01.06.2024).

23. Кузнєцов О. Л., Андрєєва Л. М. Використання нейронних мереж для класифікації новин (дата звернення:01.06.2024).

24. Левченко І. П., Гончарук Д. С. Моделювання процесів виявлення дезінформації (дата звернення:01.06.2024).

25. Литвин О. В., Тимченко І. В. Аналіз методів класифікації тексту для виявлення фейкових новин (дата звернення:01.06.2024).

26. Марченко А. С., Ковальчук Т. І. Методи виявлення фейкових новин на основі машинного навчання (дата звернення:05.06.2024).

27. Мельник В. Г., Гриценко Ю. О. Використання алгоритмів машинного навчання для обробки текстових даних (дата звернення:01.06.2024).

28. Михайленко Л. П., Кривенко І. С. Методи машинного навчання для аналізу текстової інформації (дата звернення:01.06.2024).

29. Нікітенко В. С., Соколов О. М. Методи виявлення дезінформації в текстовому контенті (дата звернення:01.05.2024).

30. Олійник В. І., Павлюк Т. М. Використання методів глибокого навчання для класифікації новин (дата звернення:01.06.2024).

31. Орлов П. В., Сидоренко І. М. Методи автоматичної класифікації тексту для виявлення фейкових новин (дата звернення:01.06.2024).

32. Павленко І. С., Богданова Н. В. Використання нейронних мереж для детекції дезінформації (дата звернення:01.06.2024).

33. Поліщук О. Г., Сидоров В. М. Методи машинного навчання для аналізу текстових даних (дата звернення:01.06.2024).

34. Попович А. С., Кравець В. П. Методи обробки тексту для виявлення фейкових новин (дата звернення:03.06.2024).

35. Пугач О. В., Тарасенко М. О. Системи автоматичного виявлення дезінформації (дата звернення:01.06.2024).

36. Рибак І. В., Чорнобай С. І. Використання методів машинного навчання для виявлення фейкових новин (дата звернення:01.06.2024).

37. Сидоренко В. С., Нікітін В. О. Методи текстової класифікації для виявлення дезінформації (дата звернення:02.06.2024).

38. Сліпченко В. І., Коваль І. П. Аналіз методів обробки текстових даних для детекції фейкових новин (дата звернення:01.06.2024).

39. Соколов А. В., Чернявська Н. М. Використання методів глибокого навчання для виявлення фейкової інформації (дата звернення:01.06.2024).

40. Степаненко О. М., Ткачук Т. В. Методи аналізу тексту для виявлення дезінформації (дата звернення:01.06.2024).

41. Тарасов В. В., Поліщук А. Г. Використання алгоритмів машинного навчання для класифікації тексту (дата звернення:01.06.2024).

42. Тимошенко Л. С., Орлов І. М. Методи обробки тексту для виявлення фейкових новин (дата звернення:03.06.2024).

43. Ткаченко О. В., Гончарук І. П. Використання нейронних мереж для класифікації текстових даних (дата звернення:01.06.2024).

44. Троян В. П., Марченко В. І. Методи машинного навчання для аналізу дезінформації (дата звернення:01.06.2024).

45. Ульянов О. Г., Бєлов В. П. Методи текстової класифікації для виявлення фейкових новин (дата звернення:03.06.2024).

46. Федоренко В. І., Коваленко О. М. Використання методів глибокого навчання для виявлення дезінформації (дата звернення:01.06.2024).

47. Фролов І. О., Васильченко Л. Г. Методи машинного навчання для аналізу тексту (дата звернення:01.06.2024).

48. Харченко В. В., Ковтун О. В. Використання алгоритмів класифікації для виявлення фейкових новин (дата звернення:04.06.2024).

49. Черненко І. С., Зінченко Т. М. Методи текстової аналітики для детекції дезінформації (дата звернення:01.06.2024).

50. Шевченко І. В., Гаврилюк Л. М. Використання методів машинного навчання для виявлення фейкових новин (дата звернення:01.06.2024).

51. Щербак А. В., Павленко Т. І. Методи обробки тексту для класифікації новин (дата звернення:05.06.2024).

52. Яковенко О. В., Сидоров І. П. Використання нейронних мереж для виявлення дезінформації у текстовому контенті (дата звернення:02.06.2024).

53. GloVe: Global Vectors for Word Representation. The Stanford Natural Language Processing Group. URL: https://nlp.stanford.edu/projects/glove/ (date of access: 06.06.2024).

54. A Step by Step Backpropagation Example. Matt Mazur. URL: https://mattmazur.com/2015/03/17/a-step-by-step-backpropagation-example/ (date of access: 06.06.2024).

55. Bloomberg - Are you a robot?. Bloomberg - Are you a robot?. URL: https://www.bloomberg.com/news/articles/2019-11-11/how-fake-news-isstoking-violence-and-anger-in-hong-kong (date of access: 06.06.2024).

56. About the Law Library | Law Library of Congress | Research Centers | Library of Congress. The Library of Congress. URL: https://www.loc.gov/research-centers/law-library-of-congress/about-this-research-center/ (date of access: 06.06.2024).