

Наталія Савіцька<sup>1</sup>, Ірина Юрчак<sup>2</sup>

<sup>1</sup> Кафедра електронних обчислювальних машин, Національний університет «Львівська політехніка», вул. С. Бандери 12, Львів, Україна, e-mail: Nataliia.Savitska.mKIKS.2022@lpnu.ua, ORCID 0009-0000-7070-2989

<sup>2</sup> Кафедра систем автоматизованого проектування, Національний університет Львівська політехніка, вул. С. Бандери 12, Львів, Україна, e-mail: iryna.y.yurchak@lpnu.ua, ORCID 0009-0005-9100-8511

## СУЧАСНІ ПІДХОДИ ДО ВИЯВЛЕННЯ ТА ПРОТИДІЇ ДЕЗІНФОРМАЦІЇ В ІНФОРМАЦІЙНИХ СИСТЕМАХ: АНАЛІЗ ТА ВДОСКОНАЛЕННЯ

Отримано: вересень 11, 2023 / Переглянуто: жовтень 02, 2023 / Прийнято: жовтень 10, 2023

© Савіцька Н., Юрчак І., 2023

<https://doi.org/>

**Анотація.** У даній науковій статті проаналізовано та охарактеризовано різні типи неправдивої інформації, що поширюються у сучасних інформаційних системах. Основна увага приділена виявленню та визначенню небезпек, які несе за собою поширення недостовірної інформації в суспільстві. Серйозні наслідки цього явища можуть сягнути глобального масштабу, тому ефективна боротьба з ним є надзвичайно важливою. Під час дослідження розглянуто різноманітні програмні методи протидії поширенню неправдивої інформації. Зокрема, проведено порівняння різних методів аналізу та фільтрації інформації, що спрямовані на виявлення та локалізацію недостовірних повідомлень. Це дозволяє визначити найбільш ефективні підходи до аналізу даних у сфері інформаційних систем та визначити оптимальні методи боротьби з поширенням неправдивих новин. Результати дослідження мають практичне застосування та можуть бути використані для вдосконалення аналізу інформації, яка надходить з різних джерел інформаційних систем. Впровадження розроблених рішень сприятиме підвищенню рівня достовірності та об'єктивності обробки інформації, покращенню якості аналізу інформації та її використанню у різних сферах суспільства.

**Keywords:** неправдива (Fake) інформація; машинне навчання (Machine Learning, ML); глибоке навчання; набір даних; довготривала короткочасна пам'ять (Long-term Short-term Memory, LSTM); згортоква нейронна мережа (Convolutional Neural Network, CNN); рекурентні нейронні мережі (Recurrent neural networks, RNN)

### Вступ

У сучасну цифрову епоху швидке поширення інформації через різні онлайн-платформи, зокрема соціальні медіа, глибоко вплинуло на те, як люди сприймають та взаємодіють з навколишнім світом. Величезний обсяг інформації, доступний для людства є безпрецедентним, причому значна її частина поширюється через соціальні мережі. Однак реальність викликає занепокоєння в тому, що значний відсоток цієї інформації є неперевіраним і вводить в оману.

Цей сплеск неперевіреної інформації має серйозні наслідки для окремих людей і суспільства в цілому. Оманлива інформація може формувати переконання, ставлення та поведінку, що часто призводить до помилкових висновків і дій. Наслідки цього можуть варіюватися від незначних непорозумінь до значного суспільного розбрату. У гіршому випадку неправдива інформація може викликати страх, спровокувати насильство або вплинути на політичні та соціальні рішення з далекосяжними наслідками.

У відповідь на цей зростаючий виклик комп'ютерні методи стали вагомим рішенням. У цих методах використовуються передові технології, зокрема машинне навчання, обробка природної мови та аналіз даних, щоб відрізнити достовірну інформацію від неправдивої. На відміну від ручної перевірки фактів, яка займає багато часу та часто є неефективною з огляду на обсяг даних, які потрібно обробити, комп'ютерні методи можуть швидко й точно оцінити достовірність інформації.

Це дослідження заглиблюється в сферу комп'ютерних методів, призначених для боротьби з дезінформацією та неправдивими новинами. Розглядається спектр складних методів від навчання з вчителем і без вчителя до навчання в ансамблі, щоб розробити ефективні моделі виявлення неправдивої інформації. Покращуючи розуміння цих методів, поставлено мету зробити внесок у поточні зусилля щодо пом'якшення згубного впливу неправдивої інформації на суспільство.

У статті представлено комплексний та аналітичний огляд сучасних комп'ютерних методів. Розглянуто їхній потенціал, сильні сторони та обмеження, виробляючи шляхи для прогресу в боротьбі з дезінформацією. Критичний аналіз та порівняння цих методів сприяє прийняттю обґрунтованих рішень для дослідників та практиків, які беруть участь у боротьбі з розповсюдженням недостовірної інформації.

### **Постановка проблеми**

Перевірка інформації на достовірність може стати для людини проблемою. Варіативність інформації та складність її формулювання ускладнюють процес ідентифікації достовірної інформації для людини. Саме через труднощі ідентифікації в інформаційні системи часто потрапляє дезінформація, чутки, міфи та інші види недостовірної інформації. Таке поширення негативно впливає на суспільство та може мати різні наслідки, від поширення неправильних знань до впливу на вибори в країнах або провокування політичних, соціальних чи військових конфліктів [1],[4].

Тому, надзвичайно важливим стало розрізнити, де саме є спотворення інформації. Існує кілька підходів до вирішення цього завдання: перевірка інформації вручну або автоматизована перевірка [6]. Розглядаючи метод ручної перевірки, можна виділити наступні переваги: людська інтуїція та досвід, складний аналіз контексту, гнучкість у розумінні нових моделей. Однак, цей підхід має ряд недоліків: витратний час і обмежена продуктивність, суб'єктивність і неточність, нездатність обробляти великі обсяги даних, залежність від досвіду.

Розглядаючи перший спосіб, можна очікувати, що інформація буде детально перевірена, але така перевірка займе багато часу і може бути неефективною через недостатню кваліфікацію особи, яка перевіряє наданий текст [3].

Розглядаючи метод автоматизованої перевірки, можна виділити наступні переваги та недоліки. Переваги: ефективність, швидкість, об'єктивність, масштабованість, систематизація та категоризація. Недоліки: неповна ефективність, обмеження інструментів, необхідність постійного оновлення, витрати на розробку методів та обслуговування.

Автоматичні системи виявлення неправдивої інформації стикаються з низкою проблем, оскільки неправдива інформація може поширюватися через багато джерел – від соціальних мереж і новинних порталів до блогів і підробних сайтів [5]. Різноманітність цих джерел ускладнює виявлення та розпізнавання неправди. Інформація дуже швидко поширюється в онлайн-середовищі, а це означає, що системи виявлення неправдивих новин повинні працювати в режимі реального часу та вдаватися до швидкого аналізу та прийняття рішень [2].

Ситуація ускладнюється ще й тим, що автори неправдивої інформації постійно адаптують та змінюють свої методи, зокрема змінюють стиль, формат, а також масштабніші спроби обійти алгоритми виявлення.

Тестові дані, які використовуються для навчання алгоритмів виявлення неправдивої інформації, також можуть бути проблемою. Набори для тестування можуть бути неоднорідними, неповними або неправильно маркованими. Це може вплинути на точність моделей і їх здатність правильно виявляти підробки [9].

Виявлення неправдивої інформації часто вимагає аналізу лінгвістичних і контекстуальних аспектів тексту. Однак, це є важким завданням через складність мови, синонімів, амфіболів та інших мовних відмінностей.

Під час аналізу вмісту для виявлення підробок у системи на основі штучного інтелекту можуть виникнути запитання щодо конфіденційності та етики. Це може обмежити обсяг дослідження та аналізу. В багатьох випадках немає чіткого «золотого стандарту» або загально визнаного визначення

недостовірної інформації, що може ускладнити створення ідеальних алгоритмів виявлення [8].

В сучасних умовах час і точність є надзвичайно важливими факторами, тому системам автоматичного виявлення недостовірної інформації частіше надаватиметься перевага над ручною перевіркою, оскільки автоматична перевірка гарантує нам, що інформацію буде перевірено швидше та ефективніше.

Основною метою даної роботи є комплексне порівняння існуючих методів і алгоритмів виявлення недостовірної інформації. Це порівняння має на меті оцінити продуктивність, точність і ефективність підходів, враховуючи їх сильні та слабкі сторони, а також їх застосовність у різних сценаріях.

Характеристика коефіцієнта пропуску неправдивої інформації буде використовуватися як показник ефективності. Тестовий набір даних, який використовується для порівняння, використовуватиметься для відтворення реалістичного вхідного сигналу, який є підбіраною колекцією новинних статей і відповідних тегів (справжніх чи фальшивих). В ідеалі він має охоплювати різноманітні теми та джерела, щоб забезпечити надійність моделі. Набір даних має бути розділений на набори для навчання та тестування (80% навчання, 20% тестування) за допомогою випадкового розподілу.

Результати дослідження слід використовувати як основу для розгляду при виборі алгоритму навчання моделей розпізнавання недостовірної інформації.

#### **Аналіз останніх досліджень**

Порівняння різних алгоритмів машинного навчання (Machine Learning, ML) і глибокого навчання (Deep Learning, DL) для виявлення неправдивих новин передбачає оцінку їх ефективності за допомогою різних показників. Однак вибір алгоритму та його продуктивність можуть відрізнятися залежно від набору даних, який використовується для тестування. Нижче наведено загальне порівняння деяких популярних алгоритмів оцінки даних.

##### *Довга короткочасна пам'ять (Long-term Short-term Memory, LSTM)*

Довга короткочасна пам'ять (LSTM) — це тип рекурентної нейронної мережі (RNN), призначеної для виявлення довготривалих залежностей у послідовності даних. Алгоритм є ефективним при обробці та аналізі текстових даних завдяки здатності зберігати інформацію в довгих послідовностях. Цю модель можна використовувати для моделювання послідовного характеру мови в новинних статтях, ефективно фіксуючи складні шаблони та залежності, які можуть вказувати на автентичність новин. Плюси: ефективний для послідовних даних, фіксує довгострокові залежності. Мінуси: вимагає інтенсивних обчислень, можливість переобладнання.

##### *Згорткова нейронна мережа (Convolutional Neural Network, CNN)*

Згорткова нейронна мережа — це архітектура глибокого навчання, яка зазвичай використовується для розпізнавання зображень, але її також можна адаптувати для аналізу тексту, розглядаючи текст як зображення. Мережа використовує згорткові шари для автоматичного й адаптивного виявлення просторових ієрархій об'єктів. У контексті тексту мережу можна застосовувати, розглядаючи слова чи n-грами як пікселі, виявляючи важливі деталі в тексті, які допомагають класифікувати новинні статті як неправдиві чи справжні. Плюси: Ефективна для текстових і графічних даних, автоматичне навчання функцій. Мінуси: потрібна велика кількість даних, чутлива до гіперпараметрів, складання інструкції щодо підготовки та подання даних.

##### *Рекурентні нейронні мережі (Recurrent neural networks, RNN)*

Рекурентні нейронні мережі — це тип штучної нейронної мережі, який підходить для послідовного аналізу даних завдяки здатності підтримувати певну форму пам'яті. У застосуванні для виявлення неправдивих новин RNN мають певні переваги та недоліки. Плюси: послідовне моделювання, пам'яті щодо попередньої інформації, довготермінові залежності, гнучкість довжини вхідних даних. Мінуси: зникаючі та розривні градієнти, вимагає інтенсивних обчислень, труднощі в охопленні глобального контексту.

### *Самоорганізуючі карти (Self-organizing map, SOM)*

Самоорганізуючі карти є особливий тип нейронних мереж з неконтрольованим навчанням. У застосуванні до виявлення неправдивих новин самоорганізуючі карти мають певні плюси та мінуси:

Плюси: навчання без контролю, топологічне відображення, зменшення розмірності, ідентифікація кластера. Мінуси: відсутність чітких міток класів, труднощі в інтерпретації, чутливість до ініціалізації та параметрів, труднощі в роботі з шумом і викидами.

### *Гібридні моделі*

Гібридні моделі поєднують кілька алгоритмів або підходів. Набувають популярності у сфері виявлення неправдивих новин завдяки їхньому потенціалу використання сильних сторін різних методів. Плюси: покращена продуктивність, різноманітне використання функцій, стійкість до змінних даних, гнучкість і можливості налаштування. Мінуси: складність у використанні, вимагає інтенсивних обчислень, збільшення зусиль з налаштування та обслуговування, складнощі з інтеграцією.

Оцінка цих технологій залежить від конкретного випадку використання та вимог завдання.

## **Результати та обговорення**

Для навчання мовних моделей потрібні великі обсяги текстових даних. Дані зазвичай збираються та позначаються мітками людьми, а потім попередньо обробляються для навчання моделі. Дані розділені на набори для навчання, перевірки та тестування. Модель навчається на навчальному наборі, а результати навченої моделі оцінюються на перевірконому наборі. Існує кілька інструментів і методів, які зазвичай використовуються для попередньої обробки даних і виділення ознак у розпізнаванні текстового контексту.

Процес навчання моделі є фундаментальним етапом. Цей етап завершується створенням функціональної моделі, яка проходить подальшу перевірку, тестування та розгортання. Ефективність моделі під час навчання є ключовою, оскільки вона безпосередньо впливає на продуктивність моделі під час інтеграції в програму для кінцевих користувачів.

На етапі навчання моделі двома критичними факторами є якість навчальних даних і вибір відповідного алгоритму. Вибір алгоритму керується конкретним випадком використання. Однак це рішення охоплює різні аспекти, включаючи складність моделі алгоритму, продуктивність, можливість інтерпретації, вимоги до обчислювальних ресурсів і швидкість. Дотримання балансу між цими факторами ускладнює процес вибору алгоритму.

Мета полягає в тому, щоб оцінити та порівняти ефективність різних моделей машинного навчання для виявлення неправдивих новин. Для порівняння обрано чотири моделі: довготривала короткочасна пам'ять (LSTM), згортова нейронна мережа (CNN), рекурентна нейронна мережа (RNN) і генеративна змагальна мережа (GAN). Оцінка охоплює типові показники: точність (Precision), запам'ятовування (Recall) та оцінка F1 (F1-score).

Для визначення даних показників введемо декілька змінних:

Правдиво-позитивний (True-positive, TP): кількість позитивних випадків, які правильно передбачили як фальшиві та насправді були фальшивими.

Правдиво-негативний (True-negative, TN): кількість негативних випадків, які правильно передбачили як істинні та насправді були справжніми.

Помилково-позитивний (False-positive, FP): кількість негативних випадків, неправильно передбачених як справжні, але насправді були помилковими.

Помилково негативний (False negative, FN): кількість позитивних випадків, які неправильно передбачили як фальшиві, але насправді були справжніми.

$$Precision = \frac{TP}{TP + FP} . \quad (1)$$

Пригадування, або чутливість, вимірює частку фактичних неправдивих новин від загальної кількості випадків, які вважаються неправдивими.

$$Recall = \frac{TR}{TP + FN} . \quad (2)$$

Оцінка F1, середньозважене значення запам'ятовування та точності, надає уявлення про загальний прогноз щодо виявлення неправдивих новин, враховуючи як помилкові позитивні, так і негативні результати.

$$F1 - score = \frac{2 \times Precision \times Recall}{Precision + Recall} . \quad (3)$$

Поширеними у сфері досліджень і програм обробки природної мови є набори даних «Fake News Detection DataSet\_English» від Kaggle.com. Ці набори даних є цінним ресурсом для навчання та тестування моделей машинного навчання спеціально для завдань класифікації тексту, що робить їх актуальними для досліджень виявлення неправдивих новин. Використання цих наборів даних дозволяє підвищити точність і ефективність моделей обробки природної мови. Це критично важливий аспект, враховуючи реальні наслідки та застосування в таких сферах, як журналістика, політика та аналіз соціальних мереж.

Ці набори даних обрано з кількох причин. По-перше, вони пропонують широкий спектр текстових даних, включаючи як неправдиві, так і справжні новини, що є фундаментальним для навчання та оцінки моделей машинного навчання для виявлення неправдивих новин. Контраст між неправдивими та правдивими новинами в наборах даних забезпечує надійне навчання моделі, допомагаючи моделі вивчити відмінні риси обох категорій.

Однак важливо визнати певні обмеження, пов'язані з цими наборами даних. Наприклад, процес маркування не завжди може бути цілком точним, вносячи потенційний шум у дані. Набори даних можуть не охоплювати весь спектр нюансів, що потенційно обмежує здатність моделі обробляти складні варіації неправдивих новин.

У дослідницькій роботі ці набори даних використовуватимуться для навчання та оцінювання різних моделей машинного навчання, включаючи, але не обмежуючись ними, довготривалу короткочасну пам'ять (LSTM), згорткову нейронну мережу (CNN), рекурентну нейронну мережу (RNN) і генеративну змагальну мережу (GAN). Велика кількість даних у наборах даних дозволить проводити глибокий аналіз продуктивності моделі на основі ключових показників, таких як точність, запам'ятовування та оцінка F1. Ця детальна оцінка допоможе отримати цінну інформацію про те, як ці моделі розрізняють неправдиві новини від справжніх, допомагаючи дослідникам у виборі найбільш ефективного підходу.

Набір даних розділено на набори для навчання та тестування (80% для навчання, 20% для тестування), кожна модель пройшла навчання на навчальному наборі та оцінювання на тестовому наборі.

У проведеному експериментальному аналізі з використанням наборів даних «Fake News Detection DataSet\_English» спостерігалася відмінна поведінка в різних моделях машинного навчання, які використовуються для виявлення неправдивих новин.

Точність, у цьому контексті, відноситься до частки правильно класифікованих екземплярів до загальної кількості розглянутих екземплярів. Це критично важливий показник, оскільки він забезпечує загальну оцінку продуктивності моделі та може використовуватися для порівняння ефективності різних моделей стандартизованим способом

$$Accuracy = \frac{NCCI}{TNI} * 100\% , \quad (4)$$

де NCCI – це кількість правильно класифікованих прикладів; TNI – це загальна кількість прикладів.

LSTM продемонстрував виняткову продуктивність у фіксуванні довгострокових залежностей у текстовому вмісті, продемонструвавши рівень точності 91,5%.

CNN продемонстрував хорошу точність у 88,2%. Однак спостерігалися труднощі під час виявлення довгострокових залежностей, що вплинуло на його здатність розпізнавати нюанси, важливі для точного виявлення неправдивих новин.

RNN також продемонструвала високу продуктивність із показником точності 90,2%.

GAN, незважаючи на інноваційність, досягнуто рівня точності 84,7%. Виникли проблеми з точки зору інтерпретації та ясності в розрізненні між неправдивими та справжніми новинами.

Результати порівняння, наведено в таблиці 1, демонструють показники точності, запам'ятовування та F1 для кожної моделі. Ці показники пропонують важливе розуміння ефективності та ефективності кожної моделі в контексті виявлення неправдивих новин.

Таблиця 1

Модель	Точність	Запам'ятовування	F1-оцінка
LSTM	0,92	0,91	0,92
CNN	0,88	0,89	0,88
RNN	0,85	0,88	0,86
GAN	0,76	0,80	0,76

Експерименти свідчать про те, що LSTM і RNN, які розроблено для обробки послідовних даних, є ефективними у виявленні шаблонів, пов'язаних з неправдивими новинами. CNN стикається з обмеженнями в обробці послідовних текстових даних. Незважаючи на те, що GAN є новою і перспективною моделлю нейронних мереж, вона потребує вдосконалення інтерпретації для практичного застосування в системах виявлення неправдивих новин. Точність є цінним показником продуктивності моделі, допомагаючи в об'єктивних порівняннях між різними моделями та їх корисності в реальних програмах.

Поєднання кількох моделей для виявлення неправдивих новин може підвищити можливості виявлення та загальну ефективність. Кожна модель має свої сильні та слабкі сторони, і їх поєднання може потенційно пом'якшити слабкі сторони та посилити сильні сторони. Нижче наведено кілька причин, чому поєднання моделей може бути корисним.

Різні моделі використовують різні підходи для аналізу текстових даних. Наприклад, LSTM зосереджується на аналізі послідовності, CNN — на локальних шаблонах, а GAN — на створенні реалістичних даних. Їх поєднання надає змогу більш повно аналізувати дані, охоплюючи різні аспекти мови та шаблонів.

Моделі можуть доповнювати одна одну, компенсуючи окремі недоліки моделі. Одна модель може бути кращою в певних контекстах або для певних типів неправдивих новин, тоді як інша може працювати краще в інших контекстах. Їх поєднання допомагає досягти більш збалансованого і точного результату.

Комбінування моделей можна розглядати як форму ансамблевого навчання, де різноманітність передбачень можна агрегувати, щоб забезпечити більш надійне та стійке передбачення. Це часто призводить до кращої продуктивності порівняно з використанням однієї моделі.

Агрегуючи прогнози з кількох моделей, можна отримати більш узагальнене рішення, яке є ефективним в більш широкому діапазоні вхідних даних. Це має вирішальне значення для змінності даних реального світу.

Об'єднання моделей може допомогти зменшити переобладнання, поширену проблему машинного навчання. Якщо одна модель ігнорує певні аспекти, інша може це компенсувати та надати більш узагальнені прогнози.

Важливо звернути увагу на можливі проблеми при використанні ансамблів:

Підвищена складність.

Обчислювальне навантаження.

Інтеграційні виклики.

Комбінування моделей є багатообіцяючим у покращенні виявлення неправдивих новин. Важливо ретельно вибирати, інтегрувати та оцінювати моделі, щоб переконатися, що переваги переважають проблеми. Добре продумана комбінація справді може використати сильні сторони кожної моделі та створити більш надійну систему виявлення неправдивих новин.

## Сучасні підходи до виявлення та протидії дезінформації в інформаційних...

У першому експерименті метою було оцінити ефективність послідовного використання довгострокової короткочасної пам'яті (LSTM) і згорткової нейронної мережі (CNN) для виявлення неправдивих даних.

LSTM демонструє помітну точність, запам'ятовування та оцінку F1, що свідчить про її вміння ідентифікувати неправдиву інформацію. Рівень точності 91,8 % додатково підтверджує загальну силу цієї моделі.

Подібним чином модель CNN демонструє високу продуктивність, зберігаючи збалансоване співвідношення точності, запам'ятовування та оцінки F1. Точність, що становить 88,2%, підкреслює здатність точно класифікувати неправдиву інформацію.

При послідовному застосуванні, починаючи з LSTM, а потім CNN, комбіновані моделі демонструють загальну продуктивність точність : 92,8%.

Результати першого експерименту наведено у таблиці 2.

Таблиця 2

Модель	Точність	Запам'ятовування	F1-оцінка
LSTM	0,92	0,91	0,92
CNN	0,88	0,89	0,88
LSTM + CNN	0,93	0,91	0,92

Застосування послідовного підходу, починаючи з LSTM і потім CNN, підвищує точність, що означає ефективну стратегію виявлення неправдивої інформації. Експеримент підкреслює ефективність кожної моделі та підкреслює потенціал для підвищення точності шляхом їх послідовного використання. Комбінований підхід оптимально використовує сильні сторони кожної моделі, потенційно створюючи більш надійну систему виявлення неправдивої інформації.

У другому експерименті мета полягала в тому, щоб оцінити продуктивність двох моделей, довгострокової короткочасної пам'яті (LSTM) і рекурентної нейронної мережі (RNN), у виявленні неправдивої інформації при послідовному використанні, починаючи з LSTM, а потім RNN.

Коли ці моделі були об'єднані в послідовний підхід, починаючи з LSTM і потім RNN, об'єднані моделі продемонстрували продуктивність щодо загальної точності: 92.6 %. Результати експерименту наведено у таблиці 3.

Таблиця 3

Модель	Точність	Запам'ятовування	F1-оцінка
LSTM	0,92	0,91	0,92
RNN	0,85	0,88	0,86
LSTM + RNN	0,93	0,90	0,92

Цей послідовний підхід призвів до підвищення точності, підкреслюючи потенціал інтеграції цих моделей для вдосконаленої системи виявлення неправдивої інформації. Комбінований підхід використовує сильні сторони кожної моделі, потенційно створюючи більш надійний механізм виявлення неправдивої інформації.

### Висновки

У сучасний період соціальні медіа набувають більшого поширення, і споживачі новин віддають перевагу цим платформам над традиційними джерелами інформації. Ця тенденція спричинила розповсюдження недостовірних повідомлень у соціальних мережах, оскільки обмін інформацією там відбувається без належної перевірки. Негативний вплив неправдивих новин посилюється, що підтверджується їхнім впливом на різні сфери життєдіяльності, підкреслюючи потенційну загрозу для суспільства.

Незважаючи на значну кількість досліджень щодо виявлення неправдивих новин, спроби класифікувати їх за допомогою спеціальних методів поки перебувають на ранніх стадіях. Однак ці методи мають великий потенціал для ефективної боротьби з неправдивими новинами. Дослідження

має на меті прояснити ці методи, стимулюючи дослідників та практиків зробити свій внесок у цю перспективну галузь.

У статті наведено результати використання різних моделей виявлення неправдивих новин. Проведено оцінку ефективності довготермінової пам'яті (LSTM), згорткової нейронної мережі (CNN), рекурентної нейронної мережі (RNN) та генеративної змагальної мережі (GAN). Досліджено комбінований підхід, який використовував ансамблі з двох моделей для максимізації переваг обох моделей. Для проведення експериментів використано набір даних "Fake News Detection DataSet\_English".

Послідовне застосування LSTM і CNN продемонструвало перспективне покращення загальної точності, яка досягла 92,8%. Цей комбінований підхід ефективно використовує переваги кожної моделі, надаючи більш надійну систему виявлення неправдивої інформації. Послідовне застосування LSTM та RNN продемонструвало продуктивність на рівні 92,6% на наборі даних.

У світлі швидкого поширення цифрової дезінформації та її значного впливу на суспільство розробка ефективних засобів боротьби з неправдивою інформацією стає надзвичайно важливою. Проведене дослідження сприяє вдосконаленню методології машинного навчання, та сприяє вирішенню проблеми. Підвищуючи точність та надійність виявлення неправдивих новин, можна зменшити потенційну шкоду та захистити людей від дезінформації.

#### **Перелік використаних джерел**

1. Karduni, A. (2019). Human-Misinformation interaction: Understanding the interdisciplinary approach needed to computationally combat false information. *ACM Digital Library*. <https://doi.org/10.1145/1122445.1122456>
2. Dennis, A., Moravec, P., & Kim, A. (2023). Search & Verify: Misinformation and source evaluations in Internet search results. *Decision Support Systems*, 171. <https://doi.org/10.1016/j.dss.2023.113976>
3. Rastogi, S., & Bansal, D. (2023). A review on fake news detection 3T's: Typology, time of detection, taxonomies. *International Journal of Information Security*. <https://doi.org/10.1007/s10207-022-00625-3>
4. BABCOCK, M., BESKOW, D., & CARLEY, K. (2018). Different Faces of False: The Spread and Curtailment of False Information in the Black Panther Twitter Discussion. *Data and Information Quality*, 11. <https://doi.org/10.1145/3339468>
5. Rohera, D., Shethna, H., Patel, K., & Thakker, U. (2022). A Taxonomy of Fake News Classification Techniques: Survey and Implementation Aspects. *IEEE Access*, 10. <https://doi.org/DOI:10.1109/ACCESS.2022.3159651>
6. Hamed, S. K., Aziz, M. J. A. A. A., & Yaakub, M. R. (2021). A review of fake news detection approaches: A critical analysis of relevant studies and highlighting key challenges associated with the dataset, feature representation, and data fusion. *Heliyon*, 9(10). <https://doi.org/10.1016/j.heliyon.2023.e20382>
7. Alghamdi, J., Lin, Y., & Luo, S. (2023). Towards COVID-19 fake news detection using transformer-based models. *Knowledge-Based Systems*, 274. <https://doi.org/10.1016/j.knosys.2023.110642>
8. Syed, L., Alsaeedi, A., Alhuri, L., & Aljohani, H. (2023). Hybrid weakly supervised learning with deep learning technique for detection of fake news from cyber propaganda. *Array*, 19. <https://doi.org/10.1016/j.array.2023.100309>
9. Mohawesh, R., Liu, X., Arini, H. M., Wu, Y., & Yin, H. (2023). Semantic graph based topic modelling framework for multilingual fake news detection. *AI Open*, 4. <https://doi.org/10.1016/j.aiopen.2023.08.004>
10. Alabaz, M., & Awajan, A. (2022). Fake-News Detection System Using Machinelearning Algorithms For Arabic-Language Content. *Journal of Theoretical and Applied Information Technology*. [Online] (Accessed: 08/10/2023)

**Nataliia Savitska<sup>1</sup>, Iryna Yurchak<sup>2</sup>**

<sup>1</sup>Department of Electronic Computing Machines, Lviv Polytechnic National University, S. Bandery str., 12, Lviv, Ukraine, E-mail: Nataliia.Savitska.mKIKS.2022@lpnu.ua, ORCID 0009-0000-7070-2989

<sup>2</sup>Computer Aided Design Systems Department, Lviv Polytechnic National University, S. Bandery street 12, Lviv, Ukraine, E-mail: iryna.y.yurchak@lpnu.ua, ORCID 0009-0005-9100-8511



## **MODERN APPROACHES TO DETECTING AND COMBATING DISINFORMATION IN INFORMATION SYSTEMS: ANALYSIS AND IMPROVEMENT**

Received: September 11, 2023 / Revised: October 02, 2023 / Accepted: October 10, 2023

© Savitska N., Yurchak I., 2023

**Abstract.** This scientific article analyzes and characterizes various types of false information disseminated in modern information systems. The main focus is on detecting and identifying the dangers associated with the spread of unreliable information in society. The serious consequences of this phenomenon can reach a global scale, making effective countermeasures critically important. The study examines and compares various software methods to counter the dissemination of false information. In particular, different methods of analyzing and filtering information aimed at detecting and localizing unreliable messages were studied and compared. This helps identify the most effective approaches to data analysis in the field of information systems and determine optimal methods to combat the spread of fake news. The conclusions of this research have practical applications and can be used to improve the analysis of information from various sources in information systems. The implementation of the solutions developed in this research will contribute to increasing the level of credibility and objectivity in information processing, thereby enhancing the quality of information analysis and its utilization in various sectors of society.

**Keywords:** fake information; machine learning (ML); deep learning; dataset; Long Short-Term Memory (LSTM); Convolutional Neural Network (CNN); Recurrent Neural