

# ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ ВИЯВЛЕННЯ ФЕЙКОВОЇ ІНФОРМАЦІЇ У СОЦМЕРЕЖАХ

Вінницький національний технічний університет

## **Анотація**

*У дослідженні запропоновано підхід до створення ІТ, що використовує штучний інтелект на основі мережі LSTM для виявлення фейкової інформації у соціальних мережах. Розробка здатна забезпечувати зручну роботу для тих, хто зацікавлений у перевірці вмісту новин, тим самим сприяючи боротьбі з поширенням дезінформації.*

**Ключові слова:** фейкова інформація; новина; штучний інтелект (ШІ); метод.

## **Abstract**

*The approach to creating IT using artificial intelligence based on the LSTM network to detect fake information in social networks is proposed in research. The technology is able to provide a convenient operation for persons who interest in checking the content of news, it encourages to the fight against the spread of misinformation.*

**Keywords:** fake; news; artificial intelligence (AI); methods.

## **Вступ**

У сучасному інформаційному просторі розповсюдження дезінформації створює серйозну проблему для цілісності поширення новин. Термін «фейкові новини» охоплює спектр оманливого контенту, від ненавмисних неточностей до зловмисних кампаній дезінформації. Вирішення цієї багатобічної проблеми вимагає інноваційних підходів, які поєднують людський інтелект із передовими ІТ.

Отже, дослідження присвячене розробленню інформаційної технології для виявлення фейкових новин із використанням можливостей штучного інтелекту (ШІ). ШІ з його здатністю до розпізнавання образів, лінгвістичного аналізу та машинного навчання стає потужним союзником у поточній боротьбі з поширенням фейкової інформації. Поява цифрових платформ демократизувала обмін інформацією, забезпечивши швидке поширення в глобальному масштабі. Однак ця демократизація також призвела до напливу дезінформації, що вимагає адаптивних рішень. Традиційні методи перевірки фактів і перевірки джерела, незважаючи на те, що вони важливі, часто стикаються з труднощами через величезний обсяг і швидкість поширення інформації.

Опрацьовано значний теоретичний доробок у напрямку досліджень, що представлений роботами таких вітчизняних та зарубіжних учених, як Кіца М., Снитюк Н., Бомчук Д., Корж О., Коровай В., Тищенко В., Наконечний В. та ін. [1–8].

## **ІТ для виявлення фейкових новин за допомогою штучного інтелекту на основі LSTM**

Автори пропонують під час розроблення ІТ для виявлення фейкових новин за допомогою штучного інтелекту на основі мережі LSTM використовувати такі етапи.

1. Парсинг веб-сторінки: це передбачає вилучення текстової інформації з різних веб-джерел.
2. Попереднє оброблення: зібрані дані проходять етапи початкового попереднього оброблення для підвищення їх якості та придатності для аналізу.
3. Інтелектуальний аналіз: для аналізу достовірності новини (вхідної інформації) використовується нейронна модель LSTM.
4. Прийняття остаточного рішення: результати аналізу ШІ надаються особі, що приймає рішення, яка остаточно оцінює достовірність інформації та заносить рішення до бази даних.
5. Навчання мережі: безперервне вдосконалення досягається шляхом періодичного навчання та тонкого налаштування мережі.

Розглянемо компоненти та процеси в структурі запропонованої ІТ для виявлення фейкових новин за допомогою штучного інтелекту на основі LSTM, що проілюстровано на рис. 1. Кожен представлений модуль системи передбачає введення призначених вхідних даних, які проходять ретельне оброблення перед переходом до наступного модуля.



Рис. 1. Модульна структура інформаційної технології визначення фейкових новин

Автори пропонують використовувати Python для програмної реалізації ІТ, оскільки вона має потужний арсенал бібліотек, що спроможні автоматизувати процес реалізації окремих модулів системи та надати доступ до потужних інструментів ІІІ.

Отже, початок процесу збирання даних починається з введення URL новинної статті. Потім, застосовується парсинг веб-сторінки для видалення відповідної інформації.

Враховуючи, що вміст веб-сторінки часто містить сторонні елементи, такі як спеціальні символи та теги HTML, етап попереднього оброблення стає обов'язковим. Щоб очистити текст шляхом видалення символів HTML, використаємо вбудовані функції Python:

```
cleanText = re.compile(r'<[^>+>|&[a-zA-Z0-9]+;')
```

У цьому модулі уточнений текст піддається численним перетворенням, включаючи видалення стоп-слів, виділення основи слів (стеммінг), видалення знаків пунктуації та нормалізацію до узгодженого регістру. Щоб видалити стоп-слова, необхідно використовувати словники стоп-слів, що реалізується засобами відповідно бібліотеки Python:

```
nltk.download('stopwords')
stop_words = set(stopwords.words('ukrainian'))
```

Для нормалізації текстового регістру в ІТ використовується методи Python:

```
lowercase_text = text.lower()
```

Для усунення пунктуації застосовано методом Python:

```
translator = str.maketrans('', '', string.punctuation)
cleaned_text = text.translate(translator)
```

Далі, необхідно перетворити оброблений вміст документа на числові вектори. Для цього застосуємо метод Word2Vec. Математичне представлення Word2Vec за допомогою косинусної подібності:

$$sim(A, B) = \cos(\theta) = (A * B) / (||A|| * ||B||),$$

де  $sim(A, B)$  – подібність косинуса,

$A$  і  $B$  – вектори слів,

$||A||$  та  $||B||$  – величини векторів.

Застосуємо бібліотеку Gensim для ефективного способу оброблення Word2Vec у Python:

```
from gensim.models import Word2Vec
from nltk.tokenize import word_tokenize
```

У модулі інтелектуального аналізу навчена мережа LSTM оцінює вміст вхідного вектору  $X$ , що представляє собою закодований масив даних аналізованої на достовірність новини. Дані, представлені у вигляді векторів слів, потрапляють до модуля нейромережі LSTM, яка генерує класифіковані дані. Класифіковані дані дозволяють LSTM оцінити достовірність новини, класифікуючи її як "Real" (достовірна) або "Fake" (фейкова). Розглянемо структуру мережі LSTM (рис. 1) для оцінки достовірності новин у соціальних мережах. Тут на вхід подаються  $X$  – вектор вхідних даних, а на виході отримуємо  $R$  – вихідний результат ("Real" або "Fake").

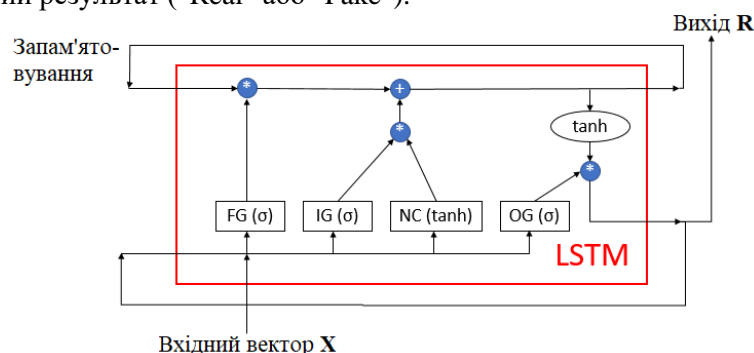


Рис. 2. Архітектура мережі LSTM

Отриманий результат передається ОПР для прийняття остаточного рішення і додавання його до бази даних із відповідним ярликом – "Real" або "Fake". Отже, цей процес є важливим у навчанні мережі LSTM, що спричинюється відповідними модифікаціями, внесеними ОПР до бази даних. Такий підхід уможливує розширення набору навчальних даних, що підвищує точність подальшої роботи ІТ, сприяючи більш надійному виявленню фейкових новин із часом.

Отже, описана ІТ здатна витягувати текстовий вміст із статей веб-новин, попередньо обробляти текст, проводити інтелектуальний аналіз і прогнозувати, надавати ОПР можливість вносити виправлення і незалежно оцінювати достовірність новин, причому, ці процеси відбуваються на тлі постійного навчання мережі за інформацією, отриманою від ОПР. На рис. 3 представлено розроблену ІТ.

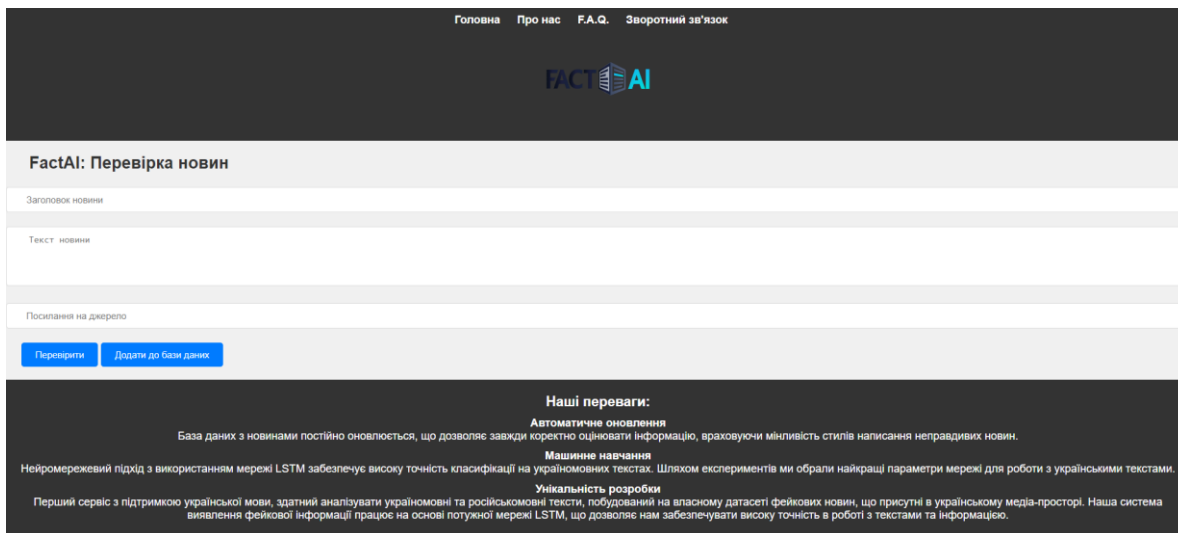


Рис. 3. Інтерфейс головної сторінки розробленої ІТ

Доведення адекватності розробленого підходу та ілюстрація роботи запропонованої ІТ. Для доведення адекватності розробленого підходу до виявлення фейкових новин було проведено аналіз результатів роботи запропонованої ІТ із результатами, що надаються офіційними джерелами інформації (рис. 4, 5).

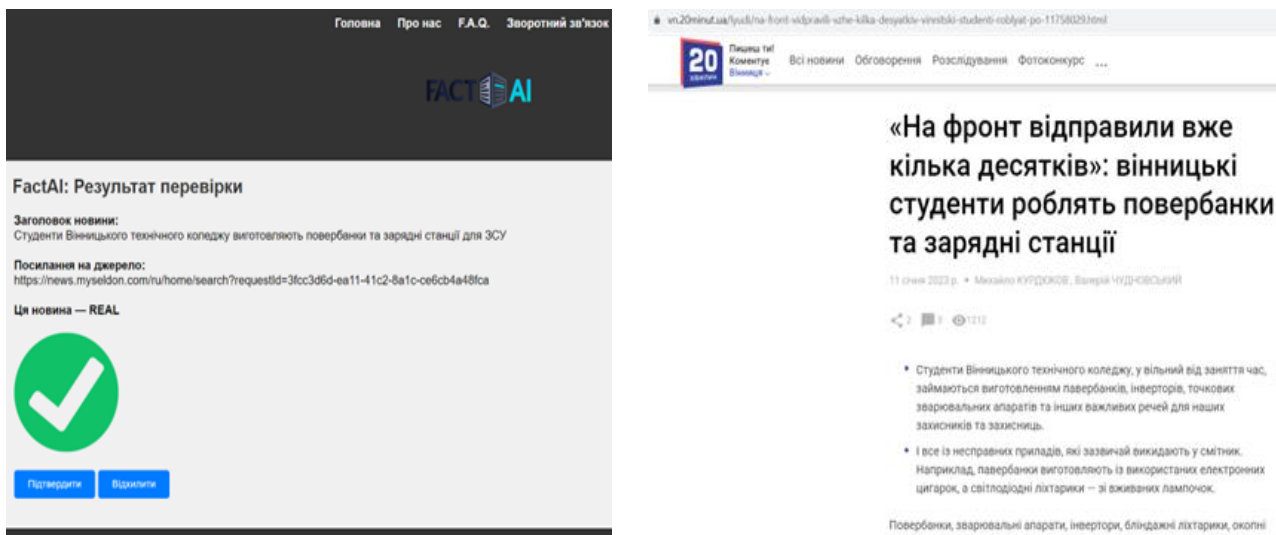


Рис. 4 – Результат перевірки новини 1 розробленою ІТ та офіційними джерелами інформації

Порівнюючи отримані результати функціонування розробленої ІТ з офіційно опублікованими даними, можна відзначити, що авторський засіб показав ідентичні результати та високу точність виявлення фейкових новин. Це свідчить про його адекватність та ефективність.

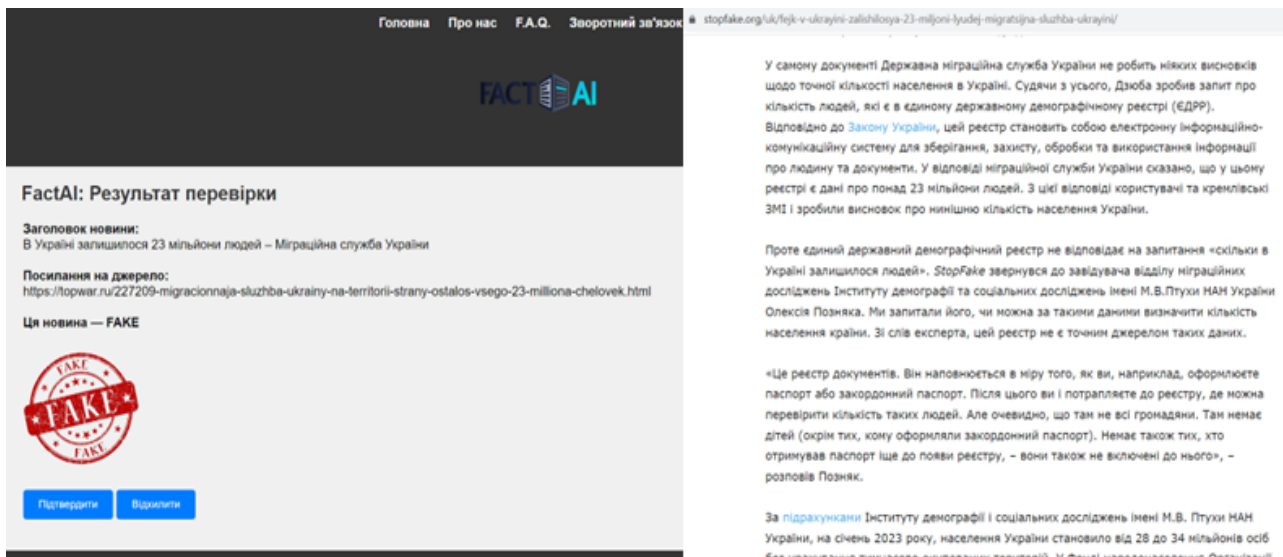


Рис. 5. Результат перевірки новини 2 розробленою ІТ офіційними джерелами інформації

**Висновки та пропозиції.** У роботі запропоновано авторський підхід до створення ІТ, що використовує штучний інтелект на основі мережі LSTM для виявлення фейкової інформації у соцмережах. Розробка здатна забезпечувати зручну роботу для тих, хто зацікавлений у перевірці вмісту новин, тим самим сприяючи боротьбі з поширенням дезінформації.

Для оцінювання ефективності та забезпечення адекватності розробленої ІТ для виявлення фейкових новин було проведено порівняльний аналіз із узгодженням його результатів з офіційними джерелами інформації. Конгруентність результатів підтверджує відповідність авторського підходу, демонструючи помітну точність, надійність і достовірність у виявленні фейкових новин.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Кіца М. Я. Особливості та методи виявлення фейкової інформації в українських ЗМІ. *Вісник університету «Львівська політехніка»*. Серія: *Журналістські науки*, 2017, 883: 1. URL: <https://science.lpnu.ua/sites/default/files/journal-paper/2019/apr/16109/kitsa.pdf> (дата звернення: 15.11.2023).
2. Сниток Н. Система виявлення фейкової інформації в мережі інтернет. *Scientific Collection «InterConf»*, 2022, 126: 201-207. URL: <https://archive.interconf.center/index.php/conference-proceeding/article/view/1373> (дата звернення: 15.11.2023).
3. Бомчук Д. В. *Фейкова інформація у соціальних медіа: сутність, виявлення, протидія*. 2023. PhD Thesis. URL: <http://ephshair.uhsp.edu.ua/handle/8989898989/7146> (дата звернення: 15.11.2023).
4. Корж О., Коровай В. Фейковий контент: види, ознаки, шляхи виявлення. *Освіта. Інноватика. Практика*, 2023, 11.7: 37-42. URL: <https://oip-journal.org/index.php/oip/article/view/216/152> (дата звернення: 15.11.2023).
5. Тищенко В., Мужанова Т. Дезінформація і фейкові новини: ознаки та методи виявлення в мережі інтернет. *Електронне фахове наукове видання «Кибербезпека: освіта, наука, техніка»*, 2022, 2.18: 175-186. URL: <https://www.csecurity.kubg.edu.ua/index.php/journal/article/view/413/341>
6. Наконечний В., Барабаш О., Лаптева Т., Міщенко А. Удосконалення методу виявлення та кластеризації джерел неправдивої інформації. *Science-based technologies*, 2022, 54.2: 105-111. URL: <https://jrn1.nau.edu.ua/index.php/SBT/article/view/16747/24063>
7. Лукова-Чуйко Н., Лаптева Т. Удосконалення методу виявлення неправдивої інформації за допомогою байєсового класифікатора. *Безпека інформації*, 2022, 28.3: 119-126. URL: <https://jrn1.nau.edu.ua/index.php/Infosecurity/article/view/17368/24692>
8. Ошийко Я. Р. *Математичне та програмне забезпечення виявлення елементів дезінформації в потоках текстових даних*. 2020. Master's Thesis. КПІ ім. Ігоря Сікорського. URL: [https://ela.kpi.ua/bitstream/123456789/39711/1/Oshyiko\\_magistr.pdf](https://ela.kpi.ua/bitstream/123456789/39711/1/Oshyiko_magistr.pdf)

**Азарова Анжеліка Олексіївна** – к.т.н., проф. каф. МБІС, Вінницький національний технічний університет, м. Вінниця.

**Пугач Володимир Сергійович** – ст. гр. 2КІТС-22м Факультету менеджменту та інформаційної безпеки, Вінницький національний технічний університет, м. Вінниця.

**Azarova Anzhelika A.** – Candidate of technical sciences, Professor of Management and security information systems department, Vinnytsia National Technical University, Vinnytsia.

**Puhach Volodymyr S.** – student of the Faculty of Management and Information Security, Vinnytsia National Technical University, Vinnytsia.