

DOI <https://doi.org/10.32689/maup.it.2026.1.1>
 УДК 004.8:004.912:004.9:316.774

АВТОМАТИЗОВАНА ПЕРЕВІРКА ТВЕРДЖЕНЬ З ВИКОРИСТАННЯМ RAG-МЕХАНІЗМУ ТА КЛАСИФІКАЦІЇ ОЗНАК

В. В. Дадиверін, О. В. Бісікало

AUTOMATED VERIFICATION OF STATEMENTS USING THE RAG MECHANISM AND SYMBOL CLASSIFICATION

Vitalii Dadyverin, Oleh Bisikalo

Анотація

Об'єктом дослідження є проблема автоматичної верифікації фактів у цифровому середовищі, насиченому дезінформацією. У роботі проаналізовано сучасні підходи до детекції фейкових новин, включаючи трансформерні архітектури, нейросемантичні та графові моделі. Додатково визначено обмеження наявних методів, зокрема популярність використання статичних ознак та погану здатність до узагальнення при постійному динамічному потоці інформації. Запропоновано власну архітектуру мультимодальної моделі, яка поєднує класифікацію стилю, детекцію AI-тексту та модуль перевірки фактів, підкріплений пошуком релевантних доказів через RAG-механізм. Результати експериментів на тестовому наборі із 1660 прикладів показали, що модель досягає високого показника Recall (84,6 %), зберігаючи прийнятну збалансованість точності (Accuracy – 78,6 %, Precision – 74,4 %, F1 – 80,8 %). Отримані результати свідчать про достатню ефективність мультизадачного навчання у системах перевірки на правдивість. Це дозволяє ефективно виявляти фейкові новини з різних джерел, хоча й із певною кількістю хибно-позитивних спрацювань, але баланс між високим Recall і нижчим Precision є обґрунтованим, оскільки система орієнтована на зменшення можливості пропуску фейкових новин. Запропонована модель є придатною для використання в реальних умовах моніторингу інформаційного простору, зокрема в контексті протидії інформаційним загрозам. Ефективність моделі пояснюється комбінацією кількох незалежних ознак (стиль, походження, фактологічність) та гнучкою системою інтеграції сигналів. Крім того, використання RAG-механізму забезпечує додатковий рівень інтерпретованості отриманих результатів з прив'язкою до зовнішніх джерел. Її можна застосовувати в онлайн-платформах з великою кількістю неструктурованих повідомлень. Підхід може бути розширений мультимедійним аналізом та адаптований для іншого визначеного мовного середовища.

Ключові слова: фактчекінг, штучний інтелект, фейкові новини, RAG, мультимодальна модель, трансформери, верифікація тверджень, дезінформація.

Abstract

The object of the study is the problem of automatic fact verification in a digital environment saturated with disinformation. The paper analyzes modern approaches to fake news detection, including transformer architectures, neurosemantic and graph models. Additionally, the limitations of existing methods are identified, in particular, the popularity of the use of static features and poor generalization ability in a constant dynamic flow of information. The author proposes his own architecture of a multimodal model that combines style classification, AI text detection and a fact-checking module, supported by the search for relevant evidence through the RAG mechanism. The results of experiments on a test set of 1660 examples showed that the model achieves a high Recall indicator (84.6 %), while maintaining an acceptable balance of accuracy (Accuracy – 78.6 %, Precision – 74.4 %, F1 – 80.8 %). The obtained results indicate sufficient effectiveness of multi-task learning in truth-checking systems. This allows for effective detection of fake news from various sources, albeit with a certain number of false positives, but the balance between high Recall and lower Precision is justified, since the system is focused on reducing the possibility of missing fake news. The proposed model is suitable for use in real-world monitoring of the information space, in particular in the context of countering information threats. The effectiveness of the model is explained by the combination of several independent features (style, origin, factuality) and a flexible signal integration system. In addition, the use of the RAG mechanism provides an additional level of interpretability of the results obtained with reference to external sources. It can be used in online platforms with a large number of unstructured messages. The approach can be expanded with multimedia analysis and adapted for another specific language environment.

Key words: fact-checking, artificial intelligence, fake news, RAG, multimodal model, transformers, claim verification, disinformation.

1. Вступ. Сьогоднішній інформаційний простір показує неймовірне зростання масштабів маніпуляцій, фейкових новин і цілеспрямованої дезінформації. Особливо яскраво це можна побачити на двох прикладах: нова адміністрація Білого дому займається маніпуляцією окремо взятими фактами для зміни картини ситуації чи розповсюджує фейки про опонентів для виправдання своїх необачних кроків, та другий приклад – це гібридна війна, яку Російська Федерація веде не лише на полі бою, а і у цифровому середовищі. Соціальні платформи, такі як Telegram, TikTok та X (Twitter), стали неперевершеними інструментами інформаційного впливу, де поширення завідомо фальшивого чи



© Дадиверін В. В., Бісікало О. В., 2026

Стаття поширюється на умовах ліцензії відкритого доступу CC BY 4.0

«перекрученого» контенту дає можливість створити іншу картинку дійсності у частини населення, що не займається подвійною перевіркою джерел задля власної інформаційної гігієни.

Такі вхідні умови вказують на збільшення потреби у системах автоматизованої перевірки фактів, які в свою чергу будуть здатні реагувати та відфільтровувати інформаційні загрози такого роду. Однак серед традиційних підходів до перевірки на факти здебільшого робиться ухил на правдивість твердження, не беручи до уваги інші важливі фактори, що можуть вказувати на маніпулятивність та штучність – маркери стилістики, генерація штучним інтелектом та джерела поширення цієї інформації.

Через наведені вище причини в межах цієї роботи запропоновано багатозадачну нейромережеву архітектуру, яка під собою об'єднує наступні задачі: визначення стилю, виявлення штучно згенерованої інформації, оцінка правдивості на основі зовнішніх джерел та об'єднання різних проміжних висновків аналізу у один спільний. Даний підхід дозволяє підвищити точність перевірки тверджень на правдивість та зробити більш стійку систему до складених маніпулятивних стратегій.

Далі в роботі наведено огляди досліджень з подібними рішеннями або проблемами, що були розв'язані в процесі розробки.

У роботі [1] наведено результати використання ансамблевих методів для обробки та аналізу тексту. Автори заявляють, що класифікатор Random Forest досягає точності 99 % на текстовому датасеті ISOT, але при використанні мультимодальних даних MediaEval 2016 ефективність вже падає до позначки 94.4 %. Також невирішеним залишаються питання переходу з узагальнених даних на реальні, такі як звичні потоки новин, та використання іншого лексичного доробку. Ньюансом, який є головною проблемою цієї методики є використання статичних ознак, що не беруть до уваги динаміку зміни маніпулятивних наративів, що властиві відповідним інформаційним просторам.

У дослідженні [2] описується розроблена гібридна модель Bi-GRU-Bi-LSTM з додатковим використанням FastText для правильної обробки арабської мови. Експериментально доводиться, що взаємодія трансформерів BERT та XLNet дає можливість досягати точності 98–99 % на текстових корпусах AFND. Проте залишилось питання з приводу правильної інтерпретації прийнятих рішень та високої ціни у вигляді обчислюваної складності для мобільних пристроїв. Ці обмеження виникли через низьку прозорість алгоритмів, бо модель видає результат без обґрунтування того, через які семантичні зв'язки чи фактичні суперечності було визначено, що ця новина може вважатись фейковою.

У роботі [3] представлено ансамблеву модель Ensemble R-X-L, що складається з трансформерів: XLNet, ALBERT та XLM-RoBERTA. Така конфігурація трансформерів дозволила досягнути точності у 98.41 % на текстових даних з платформи Kaggle. Невирішеною залишилась проблема стійкості до контекстуальних маніпуляцій, де за граматичною перевіркою проблем виявити не виходить, але зміст складається або з часткової, або з повної дезінформації. Причиною даної проблеми є максимальна сфокусованість виключно на семантичних ознаках тексту без застосування додаткової зовнішньої фактологічної перевірки.

У дослідженні [4] наводяться результати детекції реальних фейкових новин та текстів, що були згенеровані великими мовними моделями. Ансамбль двох трансформерів RoBERTA та DeBERTA показав точність 96.65 % на власному комбінованому датасеті, також автори змогли досягнути оптимізацію своєї моделі через квантизацію. Слабким місцем цього підходу є аналіз довгих текстів та глибокий семантичний розбір на контекст події. Причиною цих проблем можна вважати обмежене вікно уваги трансформерів при обробці масивного тексту, що не дозволить зауважити логічні суперечності всередині цієї новини.

У роботі [5] показується розроблена напівкерівана модель, що має в собі за основу самонавчання з механізмами інтерпретації. Описано, що використання adaptive pseudo-labeling для нерозмічених даних, дозволяє якісно розширювати базу знань, збільшуючи якість точності на 5 % в порівнянні з існуючими State-of-the-Art аналогами навіть за досить обмеженого набору анотованих прикладів. Але залишається невирішеною проблема стійкості до так званого зміщення підтвердження, коли модель підсилює невірні прогнози під час ітераційного навчання. Причиною цих проблем можна вважати відсутність зовнішнього моніторингу псевдо-міток, який може упередити можливість сприйняття хибно класифікованих одиниць даних як істинні.

У дослідженні [6] показано результат двогілкової архітектури, що має в своїй основі семантичний аналіз BERT із зовнішніми графами знань через вирівнювання триплетів. Інтеграція фактологічних доказів дозволила досягти точності розпізнавання 88.7–89.2 % на коротких текстах з соціальних мереж. Головним невирішеним питанням залишається можливість до простого масштабування системи та її можливість використовуватись для більш вузькоспеціалізованих текстових дописів. Причиною такого питання є складність ручного проектування графів знань та вартість процесу вирівнювання неструктурованого тексту для вилучення структурованих фактів.

У роботі [7] показано систему на основі RAG, яка досліджує правдивість інформації через інтернет пошук. Показано, що реалізація системи на базі GPT-4 Turbo досягає значень Micro-F1 на рівні 0.805 і це в свою чергу перевершило якісно налаштовану Llama-2, яка показує значення 0.759. Експерименти на бенчмарку RealTimeNews-25 показали 33 % зупинки аналізу через брак доказів, що показує обережність системи у випадках скудності джерел інформації. Головними невирішеними питаннями системи є часова чутливість та нормалізація відносності дат. Причиною є проблематичність трансформерів у коректному співставленні відносних часових виразів і конкретних точок у архівних справах.

Метою даного дослідження є розробка ефективного методу мультимодального текстового аналізу з визначенням правдивості інформації, що поєднує у собі наступні незалежні ознаки: стилістичні особливості текстової інформації, ймовірність згенерованості за допомогою інструментів ШІ та фактологічна достовірність, – і, це разом поєднується у одну інтегровану оцінку.

Такий підхід дозволить підвищити точність та надійність автоматизованої перевірки фактів в реальному часі, а також забезпечити якісну пояснюваність результатів для можливої подальшої перевірки чи використання користувачами.

Для досягнення мети було визначено наступні завдання:

1. Розробити модуль, що займається стилістичною класифікацією для виявлення характерних рис, притаманних маніпулятивному викладу інформації.
2. Створити модуль, що займається визначенням чи було згенеровано текст штучним інтелектом.
3. Реалізувати модуль, що буде проводити фактчекінг на основі трансформера з регресивною оцінкою подібності до зовнішніх фактів.
4. Результати роботи різних модулів інтегрувати в одну мета-оцінку, що показуватиме інтегрований висновок щодо достовірності повідомлення.
5. Провести навчання та тестування системи на репрезентативній добірці тверджень, що матиме в собі збалансовану добірку правдивих та дезінформативних тверджень.

2. Матеріали та методи досліджень. Об'єктом дослідження виступають інформаційні повідомлення або твердження з цифрового середовища, що можуть містити в собі маніпуляцію, фактологічну суперечність чи штучне походження.

Основною гіпотезою дослідження є те, що інтеграція трансформених ембедінгів із модулями детекції стилю, ідентифікації штучного походження тексту та фактологічної відповідності дає змогу підвищити достовірність автоматизованої верифікації та забезпечити високу пояснюваність фінального висновку.

Прийняті припущення:

- Вважається, що для кожного твердження у навчальному наборі підібрано відповідний текстовий доказ і цього достатньо для верифікації.
- Вважається, що для кожного вхідного інформаційного доробку достатньо знайти принаймні один релевантний текстовий доказ із корпусу знань.

Прийняті спрощення:

- Дослідження обмежено тільки текстовим аналізом без урахування мультимедійного контенту.
- Для пошуку доказів використовується заздалегідь індексований текстовий корпус, без пошуку в реальному часі в мережі Інтернет.
- Дати та хронологія не будуть нормалізуватись спеціально.

Опис модулів архітектури

Архітектура запропонованої системи (рисунок 1) побудована на основі єдиного трансформерного енкодера на базі BERT та кількома спеціалізованими модулями-класифікаторами, які виконують аналіз твердження з різних аспектів. Усі ці аспекти дозволяють моделі формувати більш глибоке та багатовимірне представлення інформації на виході, інтегруючи стилістичні, семантичні та фактологічні ознаки.

1. Трансформений енкодер (Backbone)

Це спільний енкодер, що реалізовано на базі bert-base-uncased, який використовується для обробки твердження (англ. claim) або доказу (англ. evidence). З цих вхідних даних він генерує векторне представлення з використанням CLS-пулінгу [8] і, потім, ці ембедінги передають на окремі голови (модулі) для вирішення їх підзадач.

2. Style Head

Модуль, що відповідає за класифікацію стилю твердження, тобто визначає чи присутні маніпуляції, чи текст є нейтральним. Реалізований як лінійний класифікатор з SoftMax-активацією.

3. AI Detection Head

Модуль, що відповідає за розпізнавання джерела походження твердження, тобто чи написано воно людиною, чи згенеровано AI-інструментом. Архітектурно побудований аналогічно до Style Head.

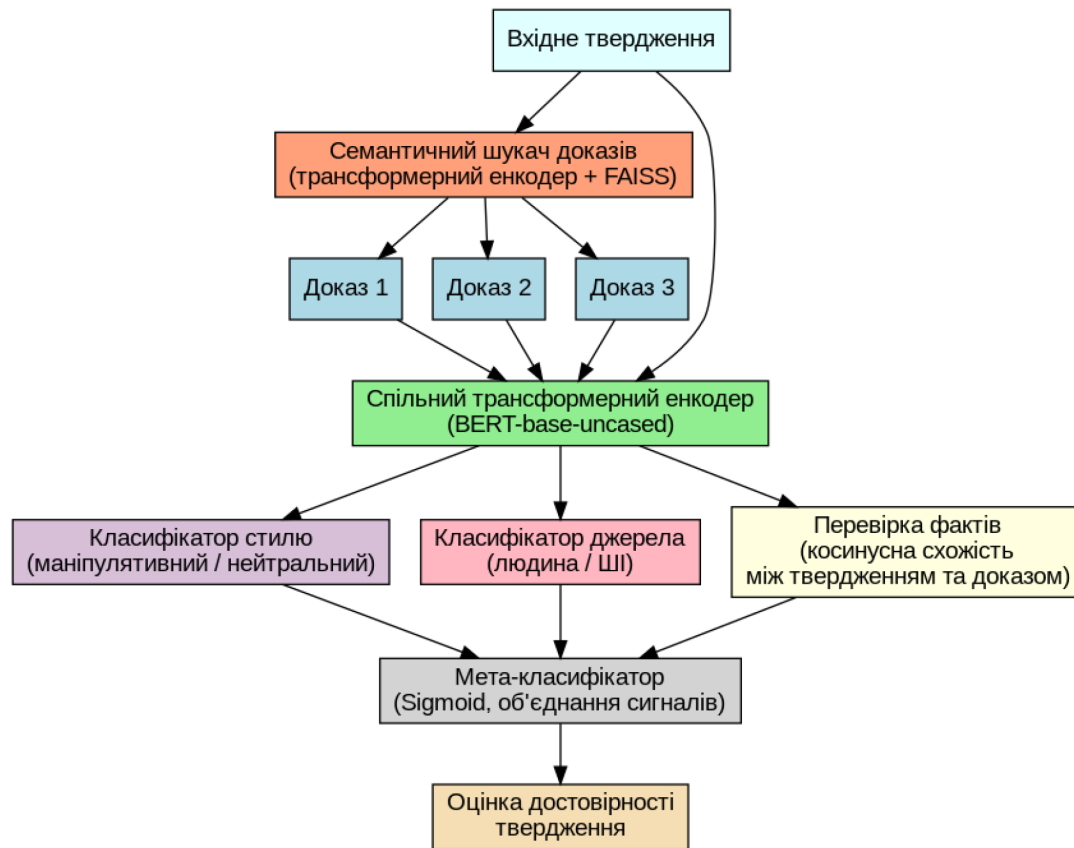


Рис. 1. Запропонована архітектура

4. Fact-Checking Head

Модуль, що відповідає за верифікацію твердження методом зіставлення векторного представлення твердження з вектором знайденого доказу. Архітектурно це виглядає як порівняння між векторними представленнями та знайденим доказом за допомогою модуля `nn.CosineSimilarity`. Реалізація відповідає концепції Retrieval-Augmented Generation [9].

5. Семантичний шукач доказів (Retriever)

Модуль, що попередньо індексує фактологічний корпус і потім виконує пошук релевантних тверджень для співставлення з вхідними даними. Модуль використовує трансформерний енкодер для кодування запиту і пошуку за схожим змістом тексту. Пошук проводиться з використанням завчасно побудованого FAISS-індексу [10].

6. Мета-класифікатор (Meta Head)

Модуль, що інтегрує вхідні сигнали з трьох інших модулів (Style, AI Detection, Fact-Checking) для формування фінального висновку щодо правдивості твердження. Архітектурно це виглядає як зв'язана нейронна мережа з активацією Sigmoid, яка на вході отримує різні оцінки і повертає ймовірність достовірності твердження. Це архітектурне рішення має підсилювати вхідні припущення і видавати узагальнений прогноз.

Корпус даних та навчальний набір

Для навчання моделі було підготовлено корпус з 1660 прикладів (рисунок 2). Кожен запис (приклад) складався з : «claim» – твердження для перевірки, «evidence» – текстовий доказ з фактологічного корпусу та анотації для різних рівнів аналізу («style_label» – нейтральний «0» чи маніпулятивний «1», «ai_label» – людина «0» чи штучний інтелект «1», «fact_label» – хибне «0» чи правдиве «1»). Багаторівневі мітки у навчальних даних дозволили використовувати підхід багатозадачного тренування (multitask learning). Такий вид навчання дозволяє оптимізувати модель під всі задачі, що у свою чергу забезпечує узгодженість роботи різних модулів в межах цієї системи.

Додатково для кожного твердження докази брались з англійськомовного підкорпусу Wikipedia [11] 2023 року, який був завчасно проіндексований бібліотекою Facebook AI Similarity Search (FAISS) [10] з плоским індексом для векторного пошуку. Це забезпечує додаткову перевірку для тверджень.

```
{
  "claim": "China was a founding member of NATO in 1949",
  "evidence": "NATO was founded by western countries. China was not a member.",
  "style_label": 1,
  "ai_label": 0,
  "fact_label": 0,
  "meta_label": 0
}
```

Рис. 2. Приклад тренувального запису

Процес навчання

Модель навчалась у багатозадачному режимі зі спільним енкдером та відокремленими модулями для кожного рівня перевірки. Для стилю твердження, його походження та збірного мета-рівня було використано крос-ентропію [12], а для перевірки на фактологічність – різниця або втрата на основі косинусної подібності [13] між векторним твердження та доказом.

Оптимізація проводилась методом AdamW [14] під час періоду 5 епох (рисунок 3), із розміром пакету – 16 і динамічним планом швидкості навчання. Гіперпараметри підбирались експериментально, беручи до уваги обмеженість вхідного корпусу прикладів.

Модель оцінювалась після кожної епохи за такими метриками: accuracy, precision, recall та F1[15]. Для вибору фінальної версії в першу чергу враховувалась F1-міра для мета-рівня, бо це є найкращим балансом між чутливістю та точністю для заключення остаточного рішення.

```
PS C:\Users\metat\Desktop\Машинна\Статті\Статті_2\project> .venv\Scripts\activate
(.venv) PS C:\Users\metat\Desktop\Машинна\Статті\Статті_2\project> python train/multitask-trainer.py --config-path config/config.yaml --exp-name base-model
Epoch 1: 100% ██████████ 55/55 [25:12:00:00, 27.50s/it]
Epoch 1 loss: 100.7767
META HEAD: {'accuracy': 0.5962732919254659, 'precision': 0.6333333333333333, 'recall': 0.6484494382022472, 'f1': 0.6368715083798083}
█ Кращу модель збережено.
Epoch 2: 100% ██████████ 55/55 [24:41:00:00, 26.93s/it]
Epoch 2 loss: 98.6576
META HEAD: {'accuracy': 0.5652173913043478, 'precision': 0.559748427672956, 'recall': 1.0, 'f1': 0.717741935483871}
█ Кращу модель збережено.
Epoch 3: 100% ██████████ 55/55 [27:29:00:00, 30.00s/it]
Epoch 3 loss: 94.0085
META HEAD: {'accuracy': 0.7329192546583851, 'precision': 0.6825396825396826, 'recall': 0.9662921348314607, 'f1': 0.8}
█ Кращу модель збережено.
Epoch 4: 100% ██████████ 55/55 [26:01:00:00, 28.39s/it]
Epoch 4 loss: 91.8844
META HEAD: {'accuracy': 0.6583850931677010, 'precision': 0.6180555555555556, 'recall': 1.0, 'f1': 0.7639484978540773}
█ Кращу модель збережено.
Epoch 5: 100% ██████████ 55/55 [26:01:00:00, 28.39s/it]
Epoch 5 loss: 80.7808
META HEAD: {'accuracy': 0.788188757763976, 'precision': 0.7391304347826086, 'recall': 0.9550561797752809, 'f1': 0.8333333333333334}
█ Кращу модель збережено.
(.venv) PS C:\Users\metat\Desktop\Машинна\Статті\Статті_2\project>
(.venv) PS C:\Users\metat\Desktop\Машинна\Статті\Статті_2\project>
```

Рис. 3. Приклад навчального процесу моделі

Інструменти та середовище

- Мова програмування: Python 3.10;
- Фреймворки та бібліотеки: PyTorch, HuggingFace Transformers, FAISS, Scikit-learn, Pandas, Json, Tqdm Python, YAML, NLTK;
- Модель: bert-base-uncased;
- Оптимізатор: AdamW;
- Середовище виконання: Локальний сервер з GPU (поточний варіант заточений під NVIDIA RTX 1050TI, але можливе налаштування на іншу відеокарту);
- Формат даних (розширення):.jsonl.

3. Результати досліджень. Оцінка ефективності проводилась на тестовому наборі з 500 прикладів (рисунок 4), де було завчасно відомо статус правдивості.

Модель продемонструвала наступні результати тестування: Accuracy ~78.6 %, Precision ~74.4 %, Recall ~84.6 % та F1 ~0.808 (80.8). Каскадна архітектура моделі дає можливість виявляти більшість хибних тверджень, дотримуючись помірного рівня точності (рисунок 5). В порівнянні

```
{
  "claim": "Joe Biden is the oldest person to assume the U.S. presidency, at age 78", "source": "en.wikipedia.org"}
{"claim": "China was a founding member of the NATO alliance in 1949", "source": "nato.int"}
{"claim": "The United Kingdom formally exited the European Union (Brexit) in January 2020", "source": "consilium.europa.eu"}
{"claim": "Switzerland is a member of the European Union", "source": "en.wikipedia.org"}
```

Рис. 4. Приклад тестових даних

з однорівневими підходами або односкладними класифікаторами, запропонована модель демонструє досить гарну чутливість (recall), що може бути критичним для задач, де головною метою є запобігання фейкам.

Нижчий рівень Precision (~74.4 %) в порівнянні з іншими моделями можна пояснити агресивною тактикою детекції: система схильна зарахувати твердження у фейки при наявності негативного

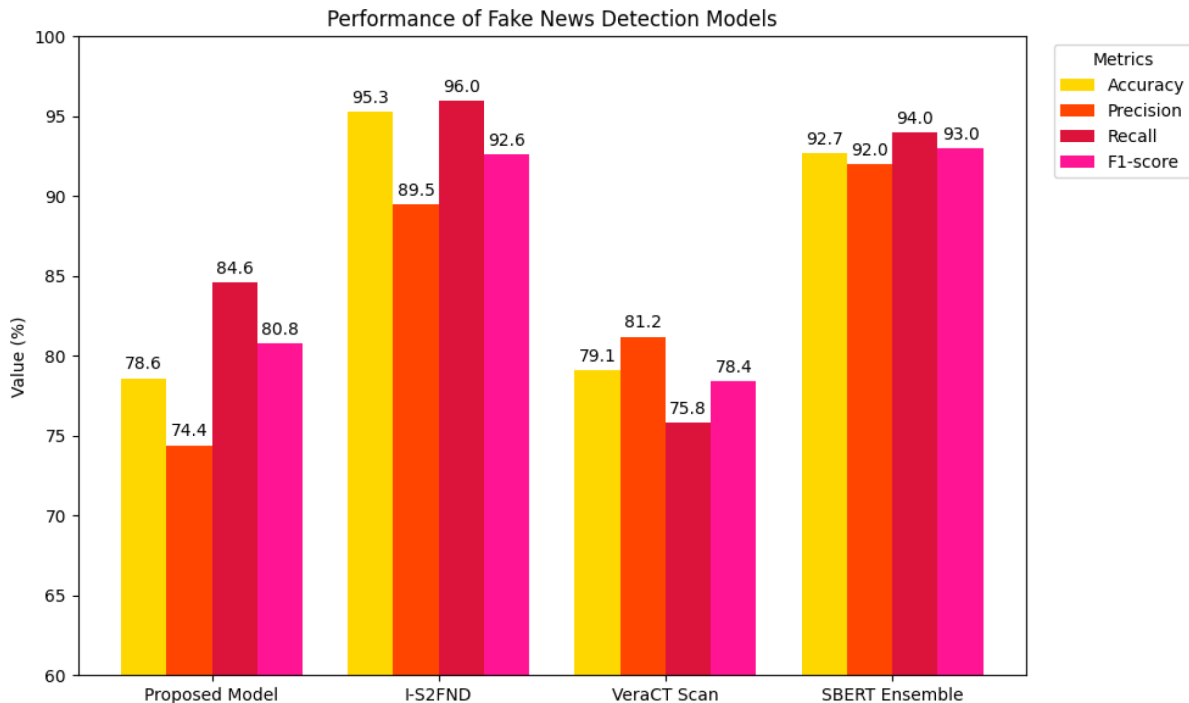


Рис. 5. Порівняння результатів тестування різних моделей

сигналу на одному з рівнів. Таке рішення зазвичай обирається для багатоступеневих систем, що орієнтовані на максимальне покриття неправдивої частини з масиву інформації. Найвище серед усіх значення Recall вказує на непогану здатність моделі знаходити приховану дезінформацію, навіть якщо інші ознаки не вказують на це. Інтегральна метрика F1 показує збалансовану якість класифікації, бо модель в цілому підтримує високий рівень правильності рішень, незважаючи на підвищену чутливість до потенційних фейків. При сценаріях, де ризик пропуску дезінформації надзвичайно важливий, такий профіль – завищений Recall та оптимальний Precision є виправданим.

Тому важливо зазначити, що каскадний класифікатор, що об'єднує в собі різні рівні аналізує, дає можливість якісніше підсилити аргументацію, а, отже, і більш впевненіше прийняти остаточне рішення з приводу правдивості вхідного твердження.

Попри зазначені вище переваги, варто розглянути та врахувати обмеження каскадної архітектури. По-перше, загальна успішність рішення значною мірою залежить від якості проходження попередніх рівнів, тобто помилка на ранньому етапі класифікації може створити ланцюгову реакцію хибних висновків на наступних етапах. По-друге, поточна версія каскадної логіки є зафіксованою, бо послідовність аналізу жорстко визначена, тобто немає можливості змінити послідовність перевірки, що позбавляє нас додаткової гнучкості та не враховує особливостей нестандартних ситуацій. По-третє, на даному етапі розробки відсутня ітеративність та темпоральність перевірки і це означає, що не враховується часовий фактор текстової добірки та не проходять цикли повторної верифікації. Ці недоліки вказують на можливі напрямки подальшого вдосконалення моделі, що мають підвищити загальну надійність представленої архітектури.

4. Висновки. Таким чином, запропонована каскадна мультимодальна модель продемонструвала достатньо високу ефективність у виявленні фейкових новин. Отримані показники (Accuracy ~78.6 %, Precision ~74.4 %, Recall ~84.6 % та F1 ~0.808) показують, що модель забезпечує достатній рівень виявлення дезінформації з оптимальним компромісом між точністю визначення та чутливістю реакції. Висока чутливість моделі до фейкової інформації є важливою перевагою для ситуацій, де пропуск одиноких проявів дезінформації має критичні наслідки, наприклад, новинні статті медичного толку, що призводять до неякісного самолікування.

Практична цінність представленої моделі полягає у можливості пристосовувати до різних платформ, що потребують виявлення фейкової інформації, адаптованість формується в першу чергу через підбір тренувальної інформації. Модульність та простота архітектури спрощує пристосування для відповідних потреб та підвищує довіру до результату, бо кожен модуль відповідає виключно за свою частину роботи і зрозумілим чином впливає на заключний результат. Попереднє твердження вказує на гарну можливість подальшого масштабування цієї архітектури, залучення нових рівнів перевірки – наприклад, модуль темпорального аналізу, що бере до уваги актуальність фактів в момент створення інформаційного доробку, або модуль ітеративного пошуку, що переформулює вхідну інформацію для повторного пошуку. Також можливе впровадження вагового мета-агрегатора, щоб комбінувати вихідні сигнали не за фіксованими правилами.

Конфлікт інтересів. Автори декларують, що не мають конфлікту інтересів стосовно даного дослідження, в тому числі фінансового, особистісного характеру, авторства чи іншого характеру, що міг би вплинути на дослідження та його результати, представлені в даній статті.

Фінансування. Дослідження проводилося без фінансової підтримки.

Доступність даних. Дані будуть надані за обґрунтованим запитом.

Використання засобів штучного інтелекту. Автори підтверджують, що не використовували технології штучного інтелекту при створенні представленої роботи.

Внесок авторів. Віталій Дадиверін: огляд існуючих рішень та підходів, створення архітектури моделі та її реалізація, навчання та тестування реалізації, написання тексту статті.

Олег Біскало: оцінка огляду літератури, пропозиції з приводу підходів до архітектури та її можливих конфігурацій, оцінка отриманих результатів тестування, внесення змін та правок у текст статті.

References:

1. Al-Shaqi, M., et al. (2024). Ensemble Techniques for Robust Fake News Detection: Integrating Transformers, Natural Language Processing, and Machine Learning. *Sensors*. Vol. 24(18). Article № 6062. DOI: <https://doi.org/10.3390/s24186062>
2. Almandouh, M., et al. (2024). Ensemble based high performance deep learning models for fake news detection. *Scientific Reports*. Vol. 14. Article № 3863. DOI: <https://doi.org/10.1038/s41598-024-76286-0>
3. Al-Ezzi, A. et al. (2022). Analysis of Deep Ensemble Transformer Model for Fake News Detection. *IEEE Access*. 2022. Vol. 10. P. 107485–107498. DOI: <https://doi.org/10.1109/ACCESS.2022.3200595>
4. Md. Ishraquzzaman et al. (2024). Ensemble Transformer-Based Detection of Fake and AI-Generated News. *Advances in Computational Intelligence and Systems*. Article ID 3268456. DOI: <https://doi.org/10.1155/acs/3268456>
5. Varshini, S. S., et al. (2023). I-S2FND: a novel interpretable self-ensembled semi-supervised model based on transformers for fake news detection. *Journal of Intelligent Information Systems*. Vol. 62. P. 233–250. DOI: <https://doi.org/10.1007/s10844-023-00821-0>
6. Yilun, Niu et al. (2024). VeraCT Scan: Retrieval-Augmented Fake News Detection with Justifiable Reasoning. arXiv preprint. arXiv:2404.01920. Available at: <https://arxiv.org/abs/2404.01920>
7. Zhao, Z., Zhou, Y., Cheng, Y. (2023). Fake News Detection Based on Knowledge-Guided Semantic Analysis. *Journal of Web Engineering*. Vol. 22(8). P. 2201–2222. DOI: <https://doi.org/10.13052/jwe1540-9589.22811>
8. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. P. 4171–4186.
9. Lewis, P., Perez, E., Piktus, A. et al. (2020). Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *Advances in Neural Information Processing Systems*. Vol. 33.
10. Facebook AI Research. Faiss: A library for efficient similarity search and clustering of dense vectors. GitHub repository. 2019–2023. Retrieved from: <https://github.com/facebookresearch/faiss>
11. Wikimedia Foundation. Корпус англomовної Вікіпедії: повне текстове дам-п-архівування статей. Wikimedia Downloads. 2023. Retrieved from: <https://dumps.wikimedia.org/enwiki/latest/>
12. Перехресна ентропія. Вікіпедія : вільна енциклопедія. 2023. Retrieved from: https://uk.wikipedia.org/wiki/Перехресна_ентропія
13. Косинус подібності. Вікіпедія : вільна енциклопедія. 2023. Retrieved from: https://uk.wikipedia.org/wiki/Косинус_подібності. Укр
14. PyTorch Contributors. torch.optim.AdamW. PyTorch Documentation. 2023. Retrieved from: <https://docs.pytorch.org/docs/stable/generated/torch.optim.AdamW.html>
15. Classification: Accuracy, Precision, Recall. Google Developers : Machine Learning Crash Course. – 2023. Retrieved from: <https://developers.google.com/machine-learning/crash-course/classification/accuracy-precision-recall>

Відомості про авторів

Англ.	Укр.
Dadyverin Vitalii Postgraduate Student Automation and Intelligent Information Technologies, Vinnytsia National Technical University vetaldadyverin@gmail.com ORCID: 0000-0001-5121-2263	Дадиверін Віталій Валерійович аспірант Автоматизації та інтелектуальних інформаційних технологій, Вінницький національний технічний університет vetaldadyverin@gmail.com ORCID: 0000-0001-5121-2263
Bisikalo Oleg Dr. Sc. (Engineering), Professor, Head of the Department of Automation and Intelligent Information Technologies, Vinnytsia National Technical University obisikalo@vntu.edu.ua ORCID: 0000-0002-7607-1943	Бісікало Олег Володимирович д-р техн. наук, професор, завідувач кафедри АІТ Автоматизації та інтелектуальних інформаційних технологій, Вінницький національний технічний університет obisikalo@vntu.edu.ua ORCID: 0000-0002-7607-1943

Дата надходження статті: 23.03.2026

Дата надходження виправленої версії статті: 10.04.2026

Дата прийняття статті: 17.04.2026

Дата публікації статті: 01.06.2026