https://doi.org/10.31891/2219-9365-2026-87-53
УДК 004.89, 004.912

МОСТОВИЙ Сергій
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»
https://orcid.org/0009-0004-9844-3981
e-mail: mostovoysergey32@gmail.com
СЕВЕРІН Андрій
Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»
https://orcid.org/0009-0009-1366-8054
e-mail: severinandrey97@gmail.com

МЕТОД ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ УКРАЇНОМОВНИХ ПОВІДОМЛЕНЬ НА ВИЯВЛЕННЯ ПРОПАГАНДИ

У роботі запропоновано метод виявлення пропаганди в коротких україномовних повідомленнях на основі поєднання TF–IDF-ознак із тональними та емоційними індикаторами. Актуальність задачі зумовлена тим, що короткі тексти характеризуються лексичною розрідженістю, неформальним стилем і високою контекстною мінливістю, що знижує ефективність суто лексичних моделей. Запропонований метод формує комбінований простір ознак, у якому TF–IDF-подання доповнюється показниками полярності, емоційної інтенсивності та експресивності повідомлення. Для класифікації використано пасивно-агресивну лінійну модель, придатну для роботи з високорозмірними розрідженими даними. Експериментальна оцінка проводилася шляхом порівняння двох варіантів: базової моделі лише з TF–IDF-ознаками та моделі з додаванням тонального блоку. Результати показали, що використання тональних ознак підвищує recall приблизно на 2,7 %, зберігаючи precision на рівні близько 80 %, а також забезпечує помірне зростання accuracy і F1-score.

Ключові слова: виявлення пропаганди, інтелектуальний аналіз тексту, обробка природної мови, аналіз тональності, україномовні повідомлення, класифікація.

MOSTOVYI Serhii, SEVERIN Andrii
National Technical University of Ukraine “Igor Sikorsky Kyiv Polytechnic Institute”

METHOD FOR INTELLIGENT ANALYSIS OF UKRAINIAN-LANGUAGE MESSAGES FOR PROPAGANDA DETECTION

This paper proposes and experimentally validates a method for binary propaganda detection in short Ukrainian language messages based on combining lexical statistical features with tonal and emotional indicators. The study addresses short text classification, where lexical sparsity, informal writing, and rapid topic shifts limit purely lexical models. The proposed approach represents each message using a TF–IDF model configured for short texts with n–gram features, and extends this representation with a compact vector of sentiment–related numeric indicators computed from a Ukrainian sentiment lexicon and expressive markers. The tonal block quantifies polarity and emotional intensity using lexicon–based positive and negative scores normalized by message length, the share of emotionally marked words, and additional indicators of expressive style such as normalized exclamation usage and capitalization ratio. The final feature space is formed by fusing the sparse TF–IDF vector with the dense tonal vector and by applying controlled weighting of feature groups to isolate the contribution of tonal information. Classification is performed with a linear Passive–Aggressive model, which is suitable for high–dimensional sparse representations and fast training. The evaluation follows a controlled comparison of two pipelines that differ only in the presence of the tonal feature block, TF–IDF only versus TF–IDF plus tonal and emotional indicators. Performance is assessed using accuracy, precision, recall, and F1–score, supported by confusion matrices and reproducible reporting. The results show that adding tonal features increases recall by about 2.7 percent on average while keeping precision around 80 percent, and yields a modest improvement in overall accuracy of about 1.4 percent and in F1 of about 1.5 percent. In practical testing scenarios, the method achieves roughly 83 percent accuracy and about 86 percent recall, supporting its applicability for monitoring information flows and flagging potentially manipulative content for subsequent expert review.

Keywords: propaganda detection, intelligent text analysis, natural language processing, sentiment analysis, Ukrainian-language messages, classification.

Стаття надійшла до редакції / Received 02.04.2026
Прийнята до друку / Accepted 12.06.2026
Опубліковано / Published 10.09.2026

ПОСТАНОВКА ПРОБЛЕМИ У ЗАГАЛЬНОМУ ВИГЛЯДІ ТА ЇЇ ЗВ’ЯЗОК ІЗ ВАЖЛИВИМИ НАУКОВИМИ ЧИ ПРАКТИЧНИМИ ЗАВДАННЯМИ

Актуальність автоматизованого виявлення пропаганди зростає в умовах цифрового інформаційного середовища, де повідомлення поширюються швидко, у великих обсягах і часто мають емоційно забарвлений характер. У контексті сучасних інформаційних протистоянь значна частина маніпулятивних наративів поширюється в соціальних медіа та месенджерах, де оперативність і масштаб ускладнюють ручну модерацію. Дослідницькі та аналітичні огляди підкреслюють еволюцію інформаційних операцій і їх адаптацію до

платформ соціальних комунікацій, що потребує прикладних засобів моніторингу та автоматичного аналізу контенту [1]. З наукового погляду завдання виявлення пропаганди є підзадачею класифікації текстів, однак ускладнюється такими чинниками:

- коротка довжина повідомлень;
- доменна мінливість і “дрейф” лексики (тобто зміна слововживання, популярних формулювань і мовних маркерів залежно від часу, теми або інформаційного середовища);
- використання емоційних тригерів, іронії, метафор та контекстних натяків.

У міжнародній практиці пропонуються як задачі детального аналізу пропагандистських технік, наприклад, фрагментна ідентифікація та класифікація типів технік, так і більш “прикладні” постановки двокласової класифікації. Зокрема, у міжнародному науковому семінарі SemEval-2020 продемонстровано, що пропаганду можна аналізувати як задачу виявлення фрагментів або технік, а сучасні рішення часто поєднують багаті ознаки та нейромережеві підходи [2].

Для україномовних даних також запропоновано низку методів машинного навчання до виявлення пропаганди та маніпулятивних технік у медіаповідомленнях [3], [4].

У даному дослідженні фокус зроблено на практично–орієнтованій двокласовій виявленні пропаганди з акцентом на:

- інтерпретоване поєднання лексико-статистичних та тонально-емоційних ознак;
- контрольоване порівняння двох постановок ознак для оцінювання внеску тональної компоненти;
- придатність до обробки потоків коротких повідомлень;
- відтворюваність експерименту за рахунок єдиного конвеєра перетворень.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

У сучасних дослідженнях виявлення пропаганди розглядається як задача автоматизованого аналізу тексту, що особливо актуалізувалася в умовах інтенсивного поширення маніпулятивного контенту в цифровому середовищі [1]. Серед наявних методів можна виокремити орієнтовані на виявлення окремих пропагандистських технік у новинних матеріалах [2], а також методи машинного навчання для класифікації україномовних текстів і медіаповідомлень [3, 4]. Окремий напрям становлять спеціалізовані набори даних і спільні задачі, присвячені виявленню маніпулятивного контенту в соціальних мережах, де враховуються особливості коротких повідомлень, неформального стилю та швидкої зміни контексту [5]. Для підвищення точності аналізу також застосовуються словники тональності та сентимент-орієнтовані ознаки, зокрема для української мови [6], а також методи, що враховують емоційне забарвлення тексту під час виявлення пропаганди [7].

Разом з тим, наявні методи мають певні обмеження. Значна частина досліджень зосереджена на новинних статтях або англомовних корпусах [2, 7], тоді як короткі україномовні повідомлення із соціальних мереж і месенджерів залишаються менш дослідженими [3–5]. Крім того, суто лексико-статистичні моделі не завжди достатньо ефективно враховують емоційну виразність, полярність, іронію та інші маркери впливу, що часто є характерними саме для коротких повідомлень [6, 7]. Тому актуальною залишається задача розроблення методу виявлення пропаганди в коротких україномовних повідомленнях, який поєднує статистичне текстове подання з тональними ознаками та придатний для практичного моніторингу інформаційних потоків.

ФОРМУЛЮВАННЯ ЦІЛЕЙ СТАТТІ

Мета дослідження — запропонувати та експериментально обґрунтувати метод виявлення пропаганди в коротких україномовних повідомленнях на основі поєднання TF–IDF і тонально-емоційних індикаторів із застосуванням лінійної пасивно-агресивної класифікації.

Завдання дослідження зводяться до наступного:

- побудувати конвеєр попередньої обробки україномовного тексту, що включає: очищення, нормалізація, токенізація, лематизація;
- сформувати два порівнювані набори ознак для подання текстових повідомлень: перший — лише на основі TF–IDF-векторизації, другий — на основі TF–IDF-векторизації, доповненої тонально-емоційними індикаторами;
- навчити та оцінити дві моделі з однаковим класифікатором щоб окремо оцінити вплив додавання тональних ознак;
- експериментально оцінити та забезпечити відтворюваність порівняння методів за поширеними метриками оцінки класифікації accuracy, precision, recall, F1.

ВИКЛАД ОСНОВНОГО МАТЕРІАЛУ ДОСЛІДЖЕННЯ

Запропонований у роботі метод реалізований як послідовний конвеєр обробки повідомлення:

1. Нормалізація повідомлення (попередня обробка, лематизація та токенізація).

2. Формування ознак:
a. формування ознак TF–IDF;
b. формування тонально-емоційних індикаторів.
3. Класифікація пасивно-агресивним алгоритмом.

Рис. 1. Схема методу інтелектуального аналізу україномовних повідомлень на виявлення пропаганди

На рис. 1 показано послідовність реалізації запропонованого методу: вхідні текстові повідомлення проходять етап попередньої обробки, після чого для них формуються TF–IDF-ознаки та тонально-емоційні індикатори. Далі ознаки об'єднуються у спільний вектор, який подається на вхід пасивно-агресивного класифікатора для визначення належності повідомлення до класу пропагандистських або непропагандистських.

Нормалізація повідомлення (попередня обробка, лематизація та токенізація). На етапі попередньої обробки застосовано нормалізацію шумів, характерних для месенджерів: URL-адреси та згадки користувачів нормалізуються до службових маркерів, пробіли уніфікуються, зберігаючи читабельний «очищений» рядок. Така нормалізація зменшує розмірність лексичного простору, підвищує узгодженість ознак між повідомленнями та знижує вплив доменно-специфічних варіацій форматування, характерних для Telegram і подібних платформ. Окремо важливо, що приведення URL і згадок до маркерів дозволяє моделі враховувати сам факт їх наявності як структурний сигнал, не прив'язуючись до конкретних рядків. Далі виконується токенізація з орієнтацією на українські літери та типові шаблони токенів (слова з апострофом, скорочення, числово-словесні поєднання), що мінімізує втрати інформації, після чого — лематизація: першочергово використовується україномовний пайплайн spaCy (моделі uk_core_news_*), а за недоступності — резервна морфологічна нормалізація. Для лематизації використано готові україномовні моделі spaCy (зокрема uk_core_news_*), призначені для токенізації, морфологічного аналізу та нормалізації українського тексту [8].

Формування ознак TF–IDF. Базовий рівень моделі використовує TF–IDF-векторизацію, яка перетворює колекцію текстів на матрицю ознак (ваг термінів), поєднуючи частоту терміна у документі та обернену документну частоту у корпусі. Для коротких повідомлень TF–IDF є доцільним базовим поданням, оскільки дозволяє виділяти слова та стійкі словосполучення, що мають підвищену дискримінативність у межах корпусу, навіть за невеликої довжини тексту. Використання біграм додатково фіксує локальні риторичні шаблони та кліше, характерні для маніпулятивного дискурсу. У реалізації TF–IDF застосовано параметри, релевантні для коротких повідомлень (зокрема, n-грамний діапазон 1–2 та відсічення надто рідкісних/надто частих термінів), а також sublinear_tf, що відповідає логарифмічному масштабуванню TF. Логарифмічне масштабування TF зменшує домінування часто повторюваних термінів у межах одного повідомлення та робить внесок лексичних ознак більш стабільним для текстів різної довжини. Загальний зміст TF–IDF і параметри векторизатора формалізовані у документації scikit-learn.

Формування тонально-емоційних індикаторів. Ключова відмінність запропонованого методу полягає в інтеграції тонально-емоційних ознак до базового TF–IDF-подання повідомлення. Для цього використовується україномовний тональний словник [6], який містить лексичні одиниці з наперед визначеними тональними характеристиками та дає змогу кількісно оцінювати емоційне забарвлення тексту. На його основі для кожного повідомлення обчислюється додатковий набір числових ознак, що відображає полярність, емоційну виразність і наявність експресивних маркерів. Таке розширення подання дає змогу врахувати не лише частотні характеристики слів, а й емоційно-тональний профіль тексту, що є важливим для виявлення маніпулятивних і пропагандистських конструкцій у коротких повідомленнях. Сформоване розширене подання використовується на етапі класифікації разом із лексико-статистичними ознаками. На рівні методу тонально-емоційний блок задається як вектор числових індикаторів, що характеризують:

- полярність повідомлення. Узагальнена оцінка позитивності/негативності, нормована за довжиною;

- суб'єктивність/емоційність. Інтенсивність емоційних маркерів, з урахуванням лексиконів та додаткових сигналів;
- нормовану частоту знаків оклику як проксі-ознаку емоційної експресії;
- частку великих літер, що часто корелює з емоційним тиском у коротких повідомленнях;
- категоріальну ознаку полярності з трьома станами негативний, нейтральний, позитивний, а також нормовану довжину як стабілізуючі характеристики.

У сукупності ці показники відображають не лише напрям полярності, а й ступінь емоційного тиску, який у коротких повідомленнях часто проявляється через експресивні маркери та нестандартне форматування. Для об'єднання лексико-статистичних і тонально-емоційних характеристик використовується формування спільного простору ознак, у якому розріджене TF–IDF-подання повідомлення доповнюється щільним вектором числових тональних індикаторів. Перед об'єднанням тонально-емоційні ознаки проходять стандартизацію для приведення їх до узгодженого масштабу, що забезпечує коректне спільне використання з TF–IDF-ознаками. Після цього обидва блоки ознак конкатенуються в єдине векторне подання повідомлення, а внесок тонально-емоційного блоку може додатково регулюватися ваговим коефіцієнтом. Така організація ознак дає змогу одночасно враховувати частотну структуру тексту та його емоційно-тональні характеристики в межах єдиної процедури класифікації. Вагове балансування блоків ознак використовується як механізм контрольованого аналізу внеску тональної компоненти, тобто дозволяє перевіряти, наскільки додаткові індикатори змінюють рішення моделі за незмінної лексичної основи. Відтворюваність такої композиції забезпечено єдиним формалізованим конвеєром перетворень, що фіксує послідовність обчислень і параметри побудови ознак.

Класифікація пасивно-агресивним алгоритмом. Фінальний класифікатор у системі — PassiveAggressiveClassifier, лінійний онлайн-метод, орієнтований на швидкі оновлення ваг і ефективну роботу з високовимірними розрідженими ознаками. Лінійна природа методу робить його придатним для задач, де ознаки мають високу розмірність і є розрідженими, як у випадку TF–IDF-подання. Це також забезпечує прозорішу поведінку моделі у порівнянні з більш складними нелінійними підходами, що є важливим для інтерпретації результатів у прикладних сценаріях. Ключові гіперпараметри включають параметр регуляризації та інтенсивності оновлення, а також обмеження ітерацій та критерій зупинки. Вибір Passive-Aggressive класифікатора зумовлений ефективністю для високовимірних розріджених ознак та придатністю до швидкого перенавчання за появи нових даних.

У виконаному дослідженні використано корпус україномовних повідомлень, наданий у межах UNLP-2025 Shared Task [5] із виявлення соціально-медійних маніпуляцій, який містить тексти з анотацією пропагандистських і маніпулятивних технік. Для цілей цього дослідження корпус приведено до задачі бінарної класифікації шляхом агрегації первинної багатокласової розмітки: повідомлення відноситься до класу «пропаганда», якщо в ньому зафіксовано щонайменше одну пропагандистську або маніпулятивну техніку; у протилежному випадку — до класу «не-пропаганда». Така постановка узгоджується з прикладним сценарієм первинної фільтрації контенту в потоці, коли системі необхідно сформувати сигнал ризику для подальшої верифікації або поглибленого аналізу людиною чи додатковими аналітичними модулями. Обране перетворення дозволяє, з одного боку, зменшити складність задачі та підвищити стабільність оцінювання на коротких повідомленнях, а з іншого — зберегти практичну інформативність розмітки, оскільки метою є фіксація самого факту наявності маніпулятивних ознак, а не їх детальне типологічне розрізнення. Додатково варто зазначити, що використання матеріалів Shared Task забезпечує відтворюваність експериментів і коректність порівняння результатів у межах спільної постановки задачі та узгоджених принципів підготовки даних. Для даних визначено ліцензійний режим, що дозволяє поширення й адаптацію за умов атрибуції та некомерційності.

Експериментальна частина побудована за принципом контрольованого порівняння двох методів, які відрізняються лише наявністю/відсутністю блоку тональних ознак. Це дозволяє інтерпретувати різницю метрик як ефект додавання тонально-емоційної компоненти.

Оцінювання якості класифікації виконано через базові метрики бінарної класифікації: accuracy, precision, recall, F1.

На рис. 2 наведено порівняння метрик для моделей "без тональності" та "з тональністю". Практичний сенс для задачі виявлення пропаганди часто пов'язаний з балансом precision/recall: з одного боку, важливо не "пропускати" потенційно небезпечні повідомлення (recall), з іншого — не перевантажувати модератора/аналітика хибними спрацюваннями (precision). Саме тому у результатах окремо акцентується приріст recall при незмінному або майже незмінному precision.

У проведеному порівнянні зафіксовано, що додавання тональних ознак підвищує повноту виявлення (Recall) у середньому на 2,7% у порівнянні з моделлю "TF–IDF only", тоді як precision зростає незначно або залишається практично незмінним (порядку ~80%). Такий профіль змін відповідає інтерпретації, що система стає чутливішою до пропагандистських повідомлень, не збільшуючи суттєво кількість хибнопозитивних спрацювань. Додатково зазначено приріст загальної точності (Accuracy) на 1,4% і F1 на 1,5% як ознаку збалансованого покращення інтегральної якості.

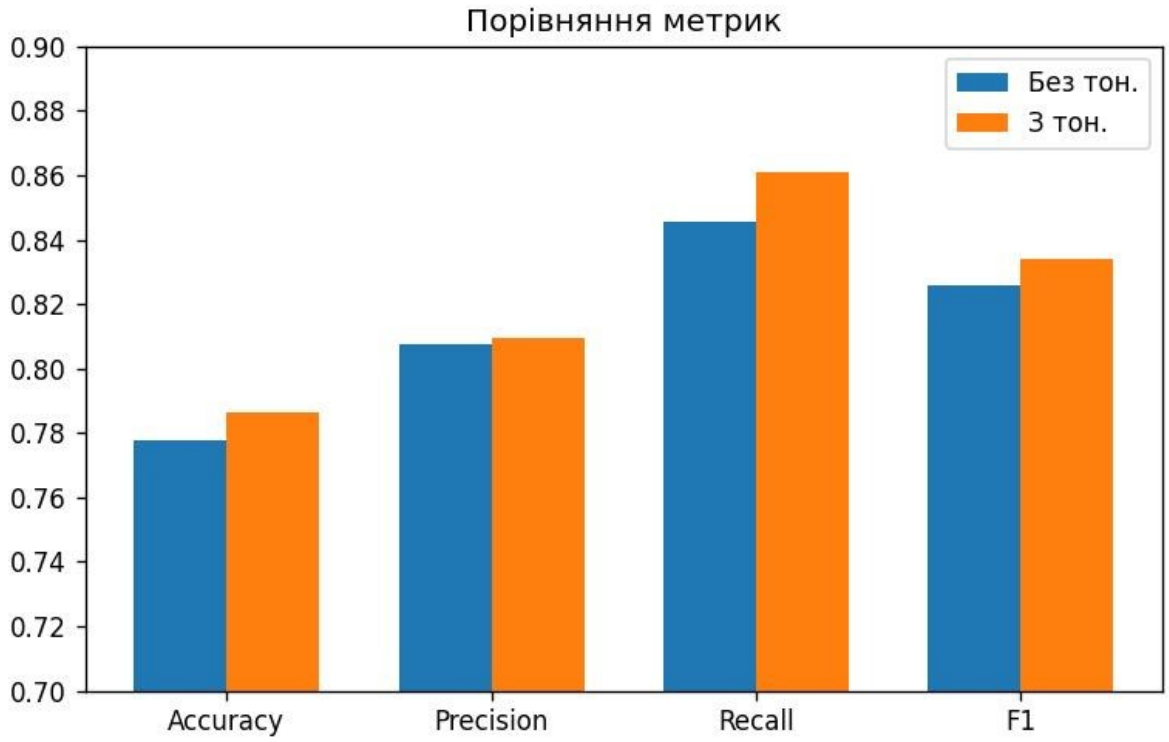

Рис. 2. Порівняння методів "без тональності" та "з тональністю"

У прикладному тестуванні метод демонструє близько 83% точності та близько 86% повноти, що узгоджується з позиціонуванням системи як придатної для моніторингу інформаційних потоків у наближеному до реального часу режимі. Узагальнено, результати підтверджують, що тонально-емоційні індикатори є корисним доповненням до суто лексико-статистичних ознак TF–IDF саме для коротких повідомлень, де емоційний тиск і лексика "настрою" можуть бути маркерами маніпулятивності [7].

ВИСНОВКИ З ДАНОГО ДОСЛІДЖЕННЯ І ПЕРСПЕКТИВИ ПОДАЛЬШИХ РОЗВІДОК У ДАНОМУ НАПРЯМІ

Результати експериментального оцінювання показали, що розширення базового TF–IDF-подання коротких україномовних повідомлень тонально-емоційними індикаторами забезпечує стабільне покращення якості виявлення пропаганди. У проведеному порівнянні зафіксовано підвищення повноти виявлення (Recall) у середньому на 2,7 % порівняно з моделлю, що використовує лише TF–IDF-ознаки, при цьому precision змінюється незначно та зберігається на рівні близько 80 %. Додатково встановлено приріст загальної точності (Accuracy) на 1,4 % та F1-міри на 1,5 %, що свідчить про збалансоване покращення інтегральної якості класифікації. У прикладному тестуванні запропонований метод демонструє близько 83 % accuracy і близько 86 % recall, що підтверджує його придатність для моніторингу інформаційних потоків у режимі, наближеному до реального часу. Загалом, отримані результати підтверджують, що тонально-емоційні індикатори є доцільним доповненням до лексико-статистичних ознак TF–IDF у задачі виявлення пропаганди саме в коротких повідомленнях, де емоційне забарвлення та маніпулятивна експресивність можуть виступати важливими маркерами впливу.

У межах виконаної роботи описано та реалізовано програмну систему для автоматизованого виявлення пропаганди в коротких україномовних повідомленнях, яка забезпечує повний цикл оброблення повідомлень — від підготовки даних і формування ознак до навчання моделі та відтворюваного оцінювання її якості. Базовою компонентою методу є TF–IDF-ознаки як статистичне подання тексту, що дозволяє фіксувати лексичні маркери й характерні словосполучення, релевантні для маніпулятивних повідомлень. Для підвищення чутливості моделі до риторичних та емоційних впливів, які часто супроводжують пропагандистські наративи, TF–IDF-подання повідомлення доповнюється набором тонально-емоційних числових ознак. Цей набір формується на основі україномовного тонального словника та додаткових експресивних маркерів і включає нормовану оцінку полярності, показник емоційної виразності, частоту знаків оклику, частку слів, написаних великими літерами, а також нормовану довжину повідомлення. Такий набір ознак не замінює лексико-статистичне подання, а доповнює його характеристиками емоційно-тонального профілю тексту, що є важливим для виявлення пропагандистського впливу в коротких повідомленнях.

Для класифікації застосовано пасивно-агресивний підхід, який ефективно працює з високовимірними розрідженими представленнями та забезпечує практичну швидкодію навчання і прогнозування. Це дозволяє

орієнтувати систему на сценарії роботи з потоками коротких повідомлень, де важливими є стабільність результату та можливість оперативного оновлення моделі при появі нових даних або зміні інформаційного контексту. Експериментальна частина роботи організована як відтворюване порівняння двох постановок — "без тональності" та "з тональністю" — за єдиним протоколом оцінювання, що включає розрахунок метрик accuracy, precision, recall і F1, аналіз матриць помилок, а також збереження та відновлення навчених моделей для повторного використання. Така методологічна рамка зменшує ризик випадкових відмінностей у налаштуваннях та підвищує коректність висновків щодо впливу тонально-емоційних індикаторів на якість виявлення пропаганди.

Практичний внесок роботи полягає у готовому модульному рішенні, придатному до інтеграції в системи моніторингу інформаційних потоків у середовищах соціальних мереж і месенджерів, де необхідно отримувати первинний сигнал ризику для подальшої верифікації або поглибленого аналізу. Запропоноване рішення дає можливість не лише застосовувати модель у прикладних задачах, а й використовувати її як базис для подальших досліджень: зокрема для розширення набору індикаторів, для уточнення правил агрегування розмітки під різні сценарії використання, а також для масштабування експериментів на більші обсяги даних і різні тематичні домени. Таким чином, отримані результати демонструють, що поєднання TF–IDF із тонально-емоційними характеристиками в межах уніфікованого процесу навчання та оцінювання є обґрунтованим напрямом підвищення якості автоматизованого виявлення пропаганди в коротких україномовних повідомленнях, а розроблена система забезпечує як методологічну цілісність експериментів, так і готовність до практичного застосування.

Література

1. Atlantic Council. Підрив України: як Росія розширила свою глобальну інформаційну війну у 2023 році [Електронний ресурс]. Atlantic Council, 2024. URL: https://www.atlanticcouncil.org/in-depth-research-reports/report/undermining-ukraine-how-russia-widened-its-global-information-war-in-2023/ (дата звернення: 16.03.2026).
2. Da San Martino G., Barrón-Cedeño A., Wachsmuth H., Petrov R., Nakov P. SemEval-2020 Task 11: виявлення пропагандистських технік у новинних статтях // Proceedings of the Fourteenth Workshop on Semantic Evaluation (SemEval-2020). 2020. P. 1377–1414. DOI: https://doi.org/10.18653/v1/2020.semeval-1.186. URL: https://aclanthology.org/2020.semeval-1.186/ (дата звернення: 16.03.2026).
3. Bazdyrev A. Виявлення дезінформації щодо російсько-української війни в підозрілих Telegram-каналах // CEUR Workshop Proceedings. 2024. Vol. 3777. ProfIT AI 2024: 4th International Workshop of IT-professionals on Artificial Intelligence, Cambridge, MA, USA, September 25–27, 2024. URL: https://ceur-ws.org/Vol-3777/short5.pdf (дата звернення: 16.03.2026).
4. Darchuk N., Zuban O., Robeiko V., Tsyhvintseva Yu., Sazhok M. Виявлення токсичності в українськомовних текстах у системі TextAttributor // CEUR Workshop Proceedings. 2024. Vol. 3909. IT&I-2024: Information Technology and Implementation, Kyiv, Ukraine, November 20–21, 2024. URL: https://ceur-ws.org/Vol-3909/Paper_6.pdf (дата звернення: 16.03.2026).
5. Kyslyi R., Romanyshyn N., Sydorskyi V. Спільне завдання UNLP 2025 з виявлення маніпуляцій у соціальних медіа // Proceedings of the Fourth Ukrainian Natural Language Processing Workshop (UNLP 2025). Association for Computational Linguistics, 2025. P. 105–111. DOI: https://doi.org/10.18653/v1/2025.unlp-1.12. URL: https://aclanthology.org/2025.unlp-1.12/ (дата звернення: 16.03.2026).
6. Lang-uk. tone-dict-uk: український словник тональності [Електронний ресурс]. URL: https://github.com/lang-uk/tone-dict-uk (дата звернення: 16.03.2026).
7. Polonijo B., Šuman S., Šimac I. Виявлення пропаганди за допомогою ансамблевого глибинного навчання з урахуванням тональності // 2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO). IEEE, 2021. P. 199–204. DOI: https://doi.org/10.23919/MIPRO52101.2021.9596654.
8. SpaCy. Українські мовні моделі [Електронний ресурс]. URL: https://spacy.io/models/uk (дата звернення: 16.03.2026).

References

1. Atlantic Council. Undermining Ukraine: How Russia Widened Its Global Information War in 2023 [Electronic resource]. Atlantic Council, 2024. URL: https://www.atlanticcouncil.org/in-depth-research-reports/report/undermining-ukraine-how-russia-widened-its-global-information-war-in-2023/ (accessed: 16.03.2026).
2. Da San Martino G., Barrón-Cedeño A., Wachsmuth H., Petrov R., Nakov P. SemEval-2020 Task 11: Detection of Propaganda Techniques in News Articles. Proceedings of the Fourteenth Workshop on Semantic Evaluation (SemEval-2020). 2020. P. 1377–1414. DOI: 10.18653/v1/2020.semeval-1.186. URL: https://aclanthology.org/2020.semeval-1.186/ (accessed: 16.03.2026).
3. Bazdyrev A. Russo-Ukrainian War Disinformation Detection in Suspicious Telegram Channels. CEUR Workshop Proceedings. 2024. Vol. 3777. ProfIT AI 2024: 4th International Workshop of IT-professionals on Artificial Intelligence, Cambridge, MA, USA, September 25–27, 2024. URL: https://ceur-ws.org/Vol-3777/short5.pdf (accessed: 16.03.2026).
4. Darchuk N., Zuban O., Robeiko V., Tsyhvintseva Yu., Sazhok M. Toxicity Detection for Ukrainian-Language Texts in the TextAttributor System. CEUR Workshop Proceedings. 2024. Vol. 3909. IT&I-2024: Information Technology and Implementation, Kyiv, Ukraine, November 20–21, 2024. URL: https://ceur-ws.org/Vol-3909/Paper_6.pdf (accessed: 16.03.2026).

5. Kyslyi R., Romanyshyn N., Sydorskyi V. The UNLP 2025 Shared Task on Detecting Social Media Manipulation. Proceedings of the Fourth Ukrainian Natural Language Processing Workshop (UNLP 2025). Association for Computational Linguistics, 2025. P. 105–111. DOI: 10.18653/v1/2025.unlp-1.12. URL: https://aclanthology.org/2025.unlp-1.12/ (accessed: 16.03.2026).

6. Lang-uk. tone-dict-uk: Ukrainian sentiment lexicon [Electronic resource]. URL: https://github.com/lang-uk/tone-dict-uk (accessed: 16.03.2026).

7. Polonijo B., Šuman S., Šimac I. Propaganda Detection Using Sentiment Aware Ensemble Deep Learning. 2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO). IEEE, 2021. P. 199–204. DOI: https://doi.org/10.23919/MIPRO52101.2021.9596654

8. SpaCy. Ukrainian language models [Electronic resource]. URL: https://spacy.io/models/uk