

УДК 004.8

Гнатюк П.В., Залуцька О.О.

*Хмельницький національний університет*

## **ПІДХІД ДО ВИЗНАЧЕННЯ ПСИХОЕМОЦІЙНОЇ ТОНАЛЬНОСТІ УКРАЇНОМОВНИХ ПОВІДОМЛЕНЬ У СОЦІАЛЬНО-ОРІЄНТОВАНИХ СЕРВІСАХ**

*Подано підхід до автоматизованого визначення психоемоційної тональності українськомовних повідомлень у соціально орієнтованих сервісах із пріоритетом прозорості та відтворюваності. Конвеєр охоплює підготовку даних, виділення ключових термінів за TF-IDF, YAKE і TextRank, зіставлення з лексиконами та навчання класифікатора з подальшим агрегуванням полярності. Ефективність оцінено на незалежних наборах за точністю, повнотою і  $F_1$  з аналізом матриць сплутувань та розбором помилок. Збереження проміжних артефактів і версій параметрів забезпечує аудит і стабільність у реальних потоках. Підхід враховує етичні вимоги щодо мінімізації персональних даних та придатний до масштабованого впровадження.*

*This work presents a method for automated detection of psycho-emotional sentiment in Ukrainian messages from socially oriented platforms with an emphasis on transparency and reproducibility. The pipeline includes data preparation, key-term extraction using TF-IDF, YAKE and TextRank, lexicon alignment, and classifier training followed by polarity aggregation. Performance is evaluated on independent sets through precision, recall and  $F_1$  together with confusion-matrix inspection and error analysis. Intermediate artifacts and parameter versions are persisted to enable full auditability and robust operation in real streams. The approach adheres to ethical constraints on personal data and is suitable for scalable deployment.*

Актуальність теми зумовлена стрімким зростанням ролі соціально орієнтованих цифрових сервісів, де короткі, часто імпульсивні повідомлення стають базовою одиницею суспільної комунікації, самоорганізації спільнот і швидкої координації дій [1, 2]. У таких потоках психоемоційна тональність виконує не другорядну, а формувальну функцію: вона впливає на довіру до джерела, модулює готовність до співпраці, спричиняє ланцюгові реакції підтримки або відторгнення та визначає видимі контури суспільних норм [3]. Для українського сегмента мережі ця проблематика має додатковий вимір через інтенсивну цифровізацію громадянської взаємодії, підвищену чутливість до ризиків дезінформації, а також потребу у відповідальній модератії без придушення легітимної критики.

Соціально орієнтовані сервіси акумулюють висловлювання різної довжини й жанру: від коротких реакцій до мікроесе з емоційними маркерами, емодзі та контекстними натяками. Психоемоційна складова в них не лише супроводжує зміст, а часто і є самим змістом, коли рішення користувача приймаються з огляду на

емоційний сигнал, а не на фактичний опис події [4]. Це вимагає коректних інструментів автоматизованого визначення тональності, здатних працювати зі стисненим контекстом, змішаними регістрами мовлення, активним код- світчингом та нестандартними графічними підсилювачами, не зводячи різноманіття виражальних засобів до грубих полярних бінарів.

Аналіз психоемоційної тональності текстів набуває дедалі більшої значущості в умовах стрімкого зростання кількості цифрової комунікації та підвищення ролі соціально-орієнтованих сервісів у взаємодії між користувачами [5]. Визначення емоційного стану автора повідомлення є важливою складовою побудови безпечного й підтримувального онлайн-середовища [6], виявлення кризових ситуацій [7], модерації контенту та адаптивної персоналізації сервісів [8, 9]. Для українського сегмента цифрових платформ це завдання ускладнюється особливостями морфології, синтаксису та стилістичних варіацій української мови [10], що потребує спеціалізованих методів обробки природної мови.

Сучасні NLP-підходи, зокрема неймережеві моделі [11], демонструють високий потенціал у класифікації емоцій [12], виявленні складних психоемоційних патернів [13] та інтерпретації контексту [14]. Використання трансформерних архітектур дає змогу опрацьовувати довгі залежності в тексті, враховувати латентні емоційні індикатори й точно розпізнавати тональність навіть у випадках іронії, сарказму та змішаних емоційних станів [15]. Значного розвитку набувають також мультимодальні підходи, що поєднують текстові дані з поведінковими або метаданими, що розширює можливості оцінки психоемоційного профілю користувача [16].

Актуальність розроблення методів визначення психоемоційної тональності українськомовних повідомлень зумовлена потребою у створенні локальних мовних моделей, здатних коректно працювати з діалектами, розмовними формами, сленгом та змішаними мовними конструкціями [17]. Наявність таких моделей є критично важливою для підвищення якості аналізу соціальних платформ, освітніх сервісів та систем психологічної підтримки, що інтегрують автоматизовані механізми оцінки стану користувачів [18].

Перспективи розвитку цього напрямку охоплюють конструювання великих українськомовних емоційно розмічених корпусів, удосконалення трансформерних моделей для роботи з низькоресурсними мовами, розроблення інтерпретованих архітектур для пояснення рішень систем та упровадження алгоритмів раннього виявлення емоційної дестабілізації. Сукупність цих напрямів формує базис для створення інтелектуальних соціальних сервісів нового покоління, здатних чутливо й точно реагувати на емоційний стан користувачів.

Українська мова у цифровому вжитку вирізняється багатшаровою варіативністю: співіснуванням літературної норми з розмовними і професійними підсистемами, міжмовними нашаруваннями, регіональними лексемами, активним словотвором і швидким оновленням сленгу. У поєднанні з суспільно значущими темами це породжує складні психоемоційні профілі, де одна й та сама поверхнева

лексема може нести протилежний емоційний знак залежно від локальної інтонації, присутності заперечення чи іронічної рамки [19]. Відповідно, підхід до автоматизованої інтерпретації повинен бути чутливим до контексту короткого висловлювання й не переносити без модифікацій припущення, вироблені на англomовних або багатоконтекстних корпусах.

Практична потреба у надійній психоемоційній аналітиці прямо пов'язана з якістю сервісів модерації, зменшенням токсичності дискусій, підтримкою ментального благополуччя користувачів та підвищенням ефективності зворотного зв'язку між громадянами, органами влади, освітніми та медичними інституціями. Помилки у визначенні тональності мають не тільки методологічну, а й соціальну ціну: вони здатні посилювати поляризацію, хибно флагувати вразливі групи, створювати хибне уявлення про консенсус або кризу, а також провокувати неадекватні управлінські рішення. Коректний інструментарій, навпаки, відкриває можливість тонкого моніторингу настроїв без втручання у свободу висловлювання та з повагою до етичних і правових обмежень.

Суттєвим викликом є характерні для соціальних платформ умови даних: нерівномірна якість текстів, висока частка повторів, кампанійні «накрутки» позитиву чи негативу, а також конвеерна поява нових меметичних форм. Без попереднього контролю якості корпусів і явної фіксації проміжних артефактів будь-яка модель, у тому числі високорівнева, демонструє нестабільність і деградацію на реальних потоках. Актуальність теми посилюється потребою прозорості та аудиту: автоматичне рішення повинно мати пояснювану структуру, щоб його можна було перевірити, відтворити і коректно оскаржити без апеляції до «чорної скриньки».

Тому, розроблений підхід включає в себе послідовний конвеєр від вхідних повідомлень до узгодженого експертного висновку, у якому кожен етап має власні артефакти й формальні точки контролю (рисунок 1). На вході перебуває масив коротких україномовних повідомлень, що надходять у первинному вигляді без ручного доопрацювання. Матеріал одразу переводиться до підсистеми підготовки, метою якої є створення стабільного представлення тексту для подальшої інтелектуальної обробки.

Перший етап охоплює нормалізацію обраного тексту, побудову мовної моделі для української та векторизацію. Нормалізація забезпечує уніфікацію графематики, усунення службових символів і підготовку до токенізації, завдяки чому знімаються технічні артефакти, що спотворюють частотний профіль. Мовна модель інтерпретує токени у відповідності до українських морфологічних і синтаксичних закономірностей, що підвищує коректність подальших вагових оцінок. Векторизація створює числове подання тексту кількома взаємодоповнюваними способами: частотно-оберненою мірою TF-IDF для стабільної корпусної специфічності, алгоритмом YAKE для відбору термінів із локальною інформативністю та TextRank для графового ранжування ключових одиниць за їхньою зв'язністю. На виході етапу формується узгоджений набір ознак, придатний і для лексиконного зіставлення, і для навчання класифікатора.

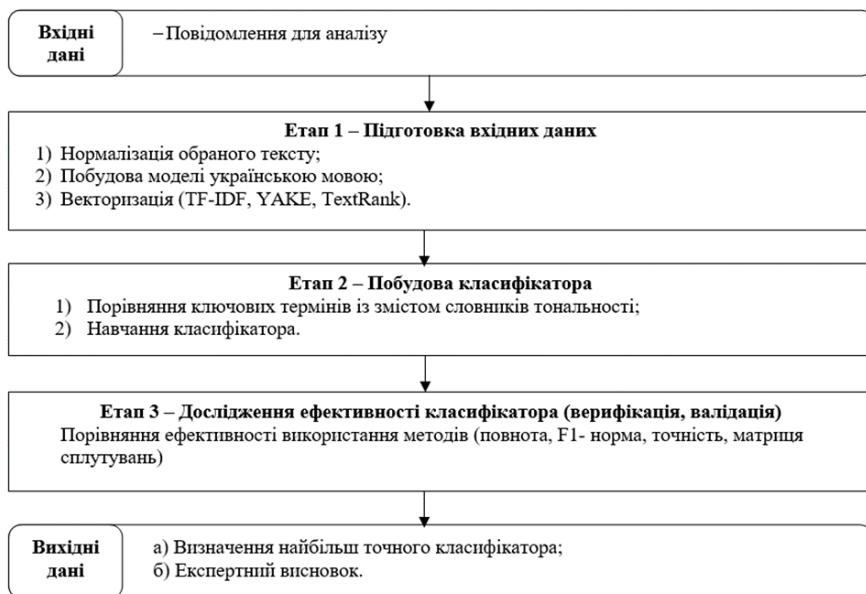


Рисунок 1 – Схема та етапи підходу

Другий етап присвячено побудові класифікатора. Спочатку відбувається порівняння виділених ключових термінів із змістом словників тональності, що дозволяє зафіксувати первинні полярні індикатори та їхні вагові внески. Далі запускається навчання класифікаційної моделі, яка поєднує інформацію з лексиконного зіставлення і векторизованих ознак. Навчання проводиться на репрезентативній вибірці з контролем дисбалансу класів і фіксацією версій параметрів. Проміжним результатом стають натреновані параметри моделі та узгоджені правила прийняття рішень, що переносяться до підсистеми оцінювання.

Третій етап реалізує дослідження ефективності класифікатора. Проводяться незалежні процедури верифікації та валідації з чітким розмежуванням наборів даних, після чого виконується порівняння обраних методів за повнотою, точністю та F1-мірою з додатковим аналізом матриць сплутувань для кожного класу. Окремо здійснюється якісний розбір хибно позитивних і хибно негативних випадків, що дозволяє коректно інтерпретувати числові показники та за потреби повертатися до попередніх етапів для уточнення параметрів нормалізації, вагових порогів або структури ознак. Завершальним продуктом є визначення найбільш точного варіанта класифікатора для заданих умов даних і сформований експертний висновок, у якому фіксуються придатність моделі до експлуатації, межі застосовності та рекомендації щодо подальшого вдосконалення.

У подальшому викладі описуються методи виокремлення ключових термінів, які надалі становитимуть основу семантичного профілю повідомлення та керуватимуть обчисленням психоемоційної тональності. Перший з них спирається на поєднання локальної ваги слова в межах одного повідомлення та його розрізняювальної здатності в масштабі всього корпусу. Інтуїція проста: термін має бути помітним у конкретному тексті, але водночас не належати до тієї частини лексики, яка зустрічається майже всюди і тому не несе змістовного сигналу. Вагове зважування поєднує ці два виміри так, щоб підсилювати слова, характерні саме для цього повідомлення, і послаблювати ті, що є фонними для всіх документів. Завдяки такій конструкції формується компактний набір лексем і стійких словосполучень, які найкраще репрезентують зміст короткого висловлювання, не втрачаючи чутливості до теми і стилю.

Другий метод зорієнтовано на внутрішньодокументні сигнали релевантності і застосовується після нормалізації тексту та сегментації на речення. Кандидатами на роль ключових одиниць стають як окремі слова, так і короткі фрази, що витримують синтаксичну цілісність. Для кожного кандидата оцінюється низка ознак, які відбивають його поведінку в конкретному документі: позиція у межах тексту, оскільки заголовкові та вступні сегменти часто містять тематичні сигнали; частота появи без штучного «накачування» повтореннями; різноманітність контекстів, що дозволяє відрізнити інформаційні носії від загальних модальних слів; розсіювання по реченнях як індикатор того, що термін підтримує зв'язність повідомлення, а не випадково виникає в одному місці. Додатково зважають на графічні індикатори, зокрема великі літери й абрєвіатури, які часто кодують власні назви або терміни домену. Після агрегування цих ознак кандидат отримує інтегральний бал, що відображає його пріоритетність саме в цьому документі.

Щоб уникнути розщеплення одного поняття на кілька технічних варіантів, результати для морфологічно споріднених форм об'єднуються. Це дає єдиний запис для леми або типової словоформи і запобігає конкуренції між близькими формами, яка спотворює ваги у коротких повідомленнях. На цьому ж кроці проводиться гармонізація написань, що відрізняються дрібними графічними деталями, але мають тотожний зміст. Так підтримується баланс між інтерпретованістю та компактністю словника ключових одиниць.

Практична інтеграція обох підходів відбувається у вигляді злиття ваг із попереднім нормуванням до спільної шкали. Частотно-інверсна схема надає глобальній дисципліни та захищає від переваги універсальних слів, натомість документно орієнтована оцінка підсилює саме ті маркери, що структурно важливі для даного тексту. У поєднанні ці сигнали формують стійкий профіль, здатний відсікти випадкові емоційні індикатори, які не підтримані контекстом, і водночас підхопити змістові осердя, що справді визначають настрої висловлювання.

У модулі визначення психоемоційної тональності цей профіль використовується як керівна маска для подальшого зіставлення зі словниковими валентностями. Високовагові терміни одержують пріоритет під час агрегування

полярності, тоді як другорядні одиниці мають обмежений вплив і не здатні самотужки змінювати знак рішення. Окремо враховуються локальні мовні явища, зокрема заперечення та інтенсифікатори, але їхній внесок модулюється саме ваговою структурою, сформованою на етапі виділення ключових термінів. Це дозволяє уникати хибно позитивних і хибно негативних рішень у випадках, коли одиничний емоційний маркер випадає з тематичної канви повідомлення.

З інженерного погляду усі проміжні артефакти обох методів зберігаються у базі даних разом із версіями параметрів, що забезпечує повний аудит і відтворюваність. У тестуванні це дає змогу порівнювати альтернативні конфігурації відбору, аналізувати чутливість до змін порогів, повертатися до будь-якого прогону та пояснювати природу розбіжностей не на рівні «чорної скриньки», а через прозорий ланцюг вагових рішень. У підсумку поєднання частотно-інверсного зважування та документно орієнтованої оцінки релевантності формує інтерпретовану і стабільну основу, на якій надалі будуються класифікаційні висновки щодо психоемоційної тональності україномовних повідомлень у соціально орієнтованих сервісах. Інтерфейс розробленого ПЗ наведено на рисунку 2.

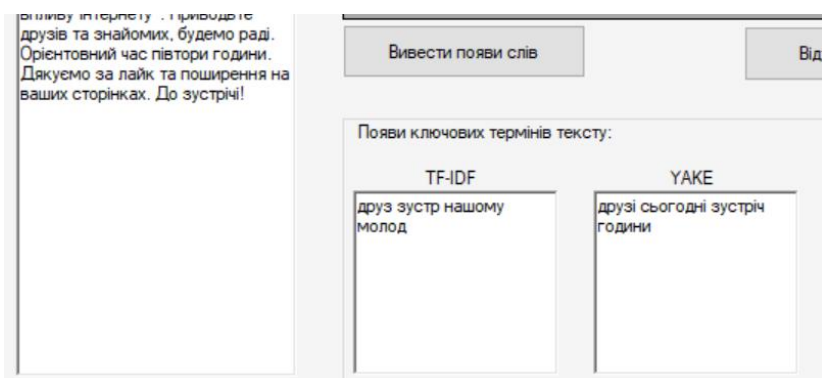


Рисунок 2 – Інтерфейс розробленого застосунку

Окремого значення набувають етичні й правові аспекти. Психоемоційні індикатори можуть трактуватися як чутлива інформація, тож технологічні рішення мають забезпечувати мінімізацію збору персональних даних, працювати з анонімізованими або агрегованими представленнями та виключати дискримінаційні ефекти. Для українського контексту, у якому цифрові платформи часто виконують функції оперативної взаємодії під час надзвичайних ситуацій та громадської мобілізації, відповідальність за інтерпретацію емоційних сигналів особливо висока.

Нарешті, завдання психоемоційного аналізу не зводиться до класифікації повідомлень на позитивні і негативні. Йдеться про побудову стійкої до збурень інфраструктури спостереження за настроями та їхніми змінами у часі, придатної до

розгортання у сервісах різної мети: від освітніх і волонтерських до муніципальних і медіаційних. Саме така рамка робить дослідження не лише науково доцільним, а й суспільно значущим, оскільки надає інструменти для раннього виявлення ескалацій, підтримки конструктивної комунікації та ухвалення рішень на підставі інтерпретованих, а не інтуїтивних емоційних сигналів.

### Перелік посилань

1. Bychkov, O., Obrusnik, T., & Dukhnovska, K. (2025). Automatic psychoemotional state analysis of text messages.
2. De Keyser, F., Dens, N., & De Pelsmacker, P. (2017). Don't be so emotional! How tone of voice and service type affect the relationship between message valence and consumer responses to WOM in social media. *Online Information Review*, 41(7), 905-920.
3. Su, L., Tang, B., & Nawijn, J. (2023). How destination social responsibility shapes resident emotional solidarity and quality of life: Moderating roles of disclosure tone and visual messaging. *Journal of Travel Research*, 62(1), 105-120.
4. Roche, J. M., Peters, B., & Dale, R. (2015). "Your tone says it all": The processing and interpretation of affective language. *Speech Communication*, 66, 47-64.
5. Гладун О.В., Мазурець О.В., Залуцька О.О. Аналіз прикладної імплементації методу нейромережевого виявлення емоційного стану людини у режимі реального часу. Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024». 15-16 листопада 2024. Хмельницький, 2024. с. 142-148.
6. Shevchuk P., Molchanova M., Mazurets O. Software for Text Messages Reliability Analysis Based on the Machine Learning Models Ensemble. Proceedings of IV International Scientific and Practical Conference «Innovative research and perspectives of the development of science and technology». January 29-31, 2024. Stockholm, Sweden. 2024. Pp. 347-354.
7. Молчанова М.О., Мазурець О.В., Собко О.В., Віт Р.В., Назаров В.В. Алгоритм виявлення аб'юзивного вмісту в україномовному аудіоконтенті для імплементації в об'єктно-орієнтовану інформаційну систему. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2024. №1 (331). С. 101-106.
8. Mazurets O., Sobko O., Vit R., Pasternak V. Practical Approach for Detection by Deep Learning of Target Objects of Subject Area Based on Semantic Connectivity Indicators in Audio Database. Proceedings of XXIV International Scientific and Practical Conference «Modern Scientific Challenges are the Driving Force of the Development of Scientific Research». May 22-24, 2024. Bruges, Belgium. International Scientific Unity. 2024. Pp. 91-96.
9. Мазурець О.В., Тимофієв І.А., Кліменко В.І., Тищенко О.О. Метод виявлення депресивного стану пов'язаного із навчанням у закладах освіти із використанням нейромережі дуальної архітектури. Науковий журнал «Вісник Херсонського національного технічного університету». 2024. №4 (91). С. 311-318.
10. Залуцька О.О., Мазурець О.В. Інформаційний портрет ключових термінів у цифрових навчальних матеріалах. Матеріали III Міжнародної науково-практичної конференції «Сучасні інформаційні технології та інноваційні методики навчання: досвід, тенденції, перспективи». Тернопіль, 2019. С.120-122.
11. Войчишин О.О., Залуцька О.О., Попов Ю.М., Купрійчук В.О. Інформаційна технологія автоматизованого формування семантичного ядра цифрових текстів. Збірник наукових праць

- за матеріалами XIII Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2021». Хмельницький, 2021. с. 298-305.
12. Молчанова М.О., Мазурець О.В., Собко О.В., Кліменко В.І., Андрощук В.І. Метод нейромережевого виявлення кібербулінгу з використанням хмарних сервісів та об'єктно-орієнтованої моделі. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2024. №2 (333). С. 200-206.
13. Залуцька О.О., Молчанова М.О., Віт Р.В., Мазурець О.В. Конфігурування нейронної мережі для класифікації емоційної тональності текстової інформації за показниками семантичної зв'язності. Збірник наукових праць за матеріалами XV Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2023». Хмельницький, 2023. с. 102-107.
14. Mazurets O.V., Sobko O.V., Molchanova M.O., Zalutska O.O., Yurchak A.V. Practical Implementation of Neural Network Method for Stress Features Detection by Social Internet Networks Posts. Global Science: Prospects and Innovations. Proceedings of the II International Scientific and Theoretical Conference «Scientific Review of the Actual Events, Achievements and Problems». May 31, 2024. Berlin, Federal Republic of Germany: International Center of Scientific Research. 2024. Pp. 160-167.
15. Blazhuk V., Mazurets O., Zalutska O. An Approach to Using the mBERT Deep Learning Neural Network Model for Identifying Emotional Components and Communication Intentions. The Impact of Scientific Research on the Development of the Modern World. Proceedings of the XLIV International scientific and practical conference. October 23-25, 2024. Dubrovnik, Croatia. 2024. Pp. 79-84.
16. Юрченко Д.Ю., Мазурець О.В., Залуцька О.О., Безпрозвана Ю.Г. Підхід до візуального пояснення результатів нейромережевого аналізу емоційної тональності повідомлень у соціальних мережах. Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024». 15-16 листопада 2024. Хмельницький, 2024. с. 565-571.
17. Денисенко Б.О., Залуцька О.О. Інтелектуальні засоби кібербезпеки в інтернет та нейромережеве виявлення дезінформації. Інформаційна, функційна і кібербезпека СКІФіК2024. Матеріали четвертої науково-технічної конференції. 29-30 листопада 2024. Харків. 2024. с. 34-35.
18. Murava V., Zalutska O., Didur V., Mazurets O. Software architecture of information system for exchanging LLM thematic prompts. Global Trends in the Development of Information Technology and Science. Proceedings IV International Scientific and Practical Conference. June 25-27, 2025. Stockholm, Sweden. Pp. 121-127.
19. Залуцька О.О. Метод автоматизованого оцінювання відповідності тональності відгуків на товари в інтернет-магазинах до їх користувачької оцінки з використанням нейромереж. Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024». 15-16 листопада 2024. Хмельницький, 2024. с. 221-236.