

робити її ефективним рішенням для енергообмежених систем. Проте такий підхід має свою специфіку – для коректної роботи моделі необхідно перетворювати її у власний формат NEF, що може призвести до втрати точності детектування порівняно з оригінальною моделлю.

### Список використаної літератури

- [1] Офіційна документація мікрокомп'ютерів Raspberry PI, [Online]. Available: <https://www.raspberrypi.org> [Accessed: October 21, 2025].
- [2] Офіційна документація мікрокомп'ютерів Nvidia, [Online]. Available: <https://www.nvidia.com/en-us/autonomous-machines/embedded-systems/> [Accessed: October 21, 2025].
- [3] Офіційна документація моделей YOLO, ultralytics, [Online]. Available: <https://docs.ultralytics.com/> [Accessed: October 21, 2025].
- [4] Офіційна документація апаратного прискорювача HAILO model zoo, [Online]. Available: [https://github.com/hailo-ai/hailo\\_model\\_zoo/blob/master/docs/public\\_models/](https://github.com/hailo-ai/hailo_model_zoo/blob/master/docs/public_models/) [Accessed: October 21, 2025].

УДК: 004.8

## ПІДХІД ДО ВІЗУАЛЬНОЇ ІНТЕРПРЕТАЦІЇ РЕЗУЛЬТАТІВ АНАЛІЗУ ЕМОЦІЙНИХ ПАТЕРНІВ МАНІПУЛЯТИВНИХ ТЕХНІК У ЦИФРОВИХ ТЕКСТАХ ЗАСОБАМИ ОБРОБКИ ПРИРОДНОЇ МОВИ

Юрченко Д. Ю., Мазурець О.В.

(di4iker@gmail.com, exe.chong@gmail.com)

Хмельницький національний університет (Україна)

*У статті представлено науково обґрунтований підхід до візуальної інтерпретації результатів аналізу емоційних патернів маніпулятивних технік у цифрових текстах засобами обробки природної мови. Дослідження спрямоване на подолання проблеми непрозорості нейромережесевих моделей, які, попри високу точність класифікації, залишаються складними для пояснення та інтерпретації. Розроблено метод, що поєднує глибоку гібридну архітектуру типу CNN–BiLSTM із модельно-агностичним механізмом пояснення LIME, який забезпечує локальне відображення вагових внесків лексичних одиниць у формування емоційної оцінки тексту. Конволюційні шари виділяють локальні емоційно насичені фрагменти, а рекурентна компонента двонаправленого типу моделює динаміку емоційних станів, що дозволяє простежити розвиток маніпулятивного тону в межах повідомлення. Запропонований підхід реалізовано у вигляді візуальної карти емоційних тригерів, яка відображає причинно-наслідкові зв'язки між мовними структурами та емоційним впливом. Емпіричні результати свідчать, що поєднання CNN та BiLSTM забезпечує точність класифікації понад 90 %, а інтеграція пояснюваних моделей дає змогу інтерпретувати механізми емоційного впливу з мінімальним ризиком помилкових висновків. Застосування розробленої системи дає можливість не лише фіксувати факт наявності маніпуляції, а й виявляти її емоційно-когнітивну структуру, що підвищує рівень довіри до штучного інтелекту в аналізі цифрового контенту. Результати дослідження мають як теоретичне, так і практичне значення: вони розширюють межі пояснюваного штучного інтелекту в контексті обробки природної мови та формують основу для створення когнітивно орієнтованих систем цифрової безпеки й медіагієни..*

Проблематика аналізу маніпулятивних текстів у цифровому середовищі пов'язана з необхідністю ідентифікації латентних емоційних патернів, які використовуються для формування прихованого когнітивного впливу на користувача [1]. У сучасних комунікаційних системах, що функціонують на основі соціальних мереж, мікроблогів та інтерактивних платформ, поширення інформації супроводжується навмисними або несвідомими спробами модифікувати емоційний стан аудиторії [2]. Такі процеси відбуваються переважно через використання лексико-семантичних засобів, інтонаційних моделей, когнітивних тригерів і дискурсивних стратегій, спрямованих на виклик емоцій напруги, страху, провини або співчуття [3]. Виявлення цих

закономірностей потребує складного аналізу, який виходить за межі традиційної лінгвістичної інтерпретації та вимагає застосування методів штучного інтелекту, здатних моделювати приховану структуру емоційного забарвлення тексту.

Сучасні підходи до аналізу емоційної тональності базуються на використанні обробки природної мови, що забезпечує автоматизовану класифікацію текстів за їх емоційною спрямованістю [4]. Проте навіть високоточні нейромеревеві моделі залишаються непрозорими у поясненні прийнятих рішень, що створює ефект «чорної скриньки» та унеможливорює повне розуміння причин класифікації конкретного повідомлення. У сфері дослідження маніпулятивних технік це має особливе значення, адже не менш важливо знати не тільки, що повідомлення є маніпулятивним, а й завдяки яким саме емоційним механізмам воно справляє вплив. Для вирішення цього завдання доцільним є використання поєднання глибоких нейронних архітектур із пояснюваними моделями штучного інтелекту, здатними відобразити внесок кожного мовного елемента в остаточне рішення.

Запропонований у дослідженні підхід ґрунтується на інтеграції гібридної нейромеревевої архітектури, яка поєднує властивості згорткових і рекурентних компонентів, із модельно-агностичними механізмами пояснення LIME. Конволюційні шари забезпечують виявлення локальних ознак емоційного збудження, що виявляються у вигляді ключових фраз або лексем із високою афективною насиченістю. Рекурентна компонента двонаправленого типу BiLSTM моделює довготривалі залежності між словами, уможливаючи виявлення емоційної динаміки тексту. Поєднання цих двох механізмів створює гнучку архітектуру, яка спроможна не лише фіксувати локальні патерни, а й відтворювати їхній контекстуальний розвиток у межах повідомлення. На наступному етапі застосовується модель пояснення, що локально апроксимує поведінку нейронної мережі в околі конкретного прикладу. Така інтерпретація подається у вигляді візуального зображення, де кожне слово або фраза позначені кольоровими маркерами відповідно до їхнього впливу на результат класифікації. Це дозволяє створити візуальну карту емоційних тригерів, на підставі якої можливо зрозуміти не лише спрямованість, а й механізм маніпулятивного впливу.

Застосування запропонованої моделі (рисунок 1) продемонструвало високу ефективність у розпізнаванні складних емоційних конструкцій, що використовуються для формування маніпулятивного тону. Емпіричні результати підтвердили, що поєднання архітектур CNN та BiLSTM забезпечує точність класифікації понад дев'яносто відсотків, а інтеграція пояснюваних моделей дозволяє інтерпретувати внесок кожного токена у прогноз, мінімізуючи ризики помилкових висновків.

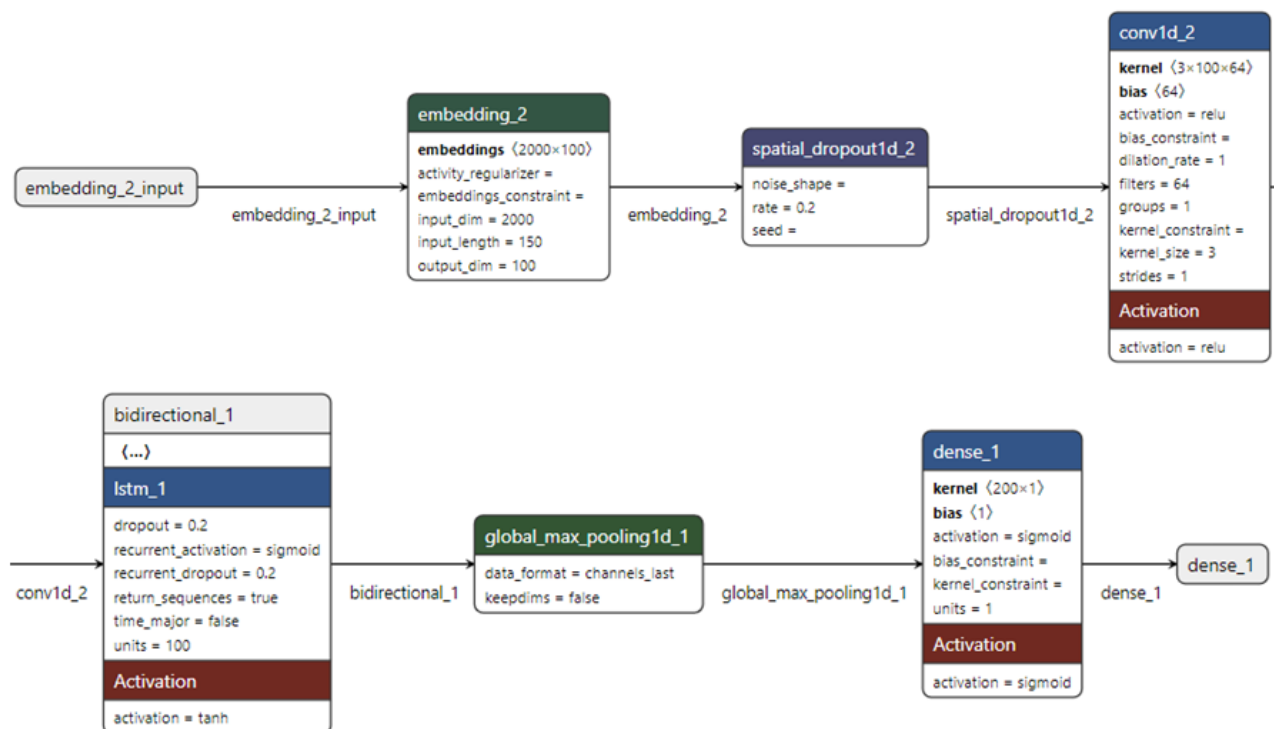


Рисунок 1 – Архітектура нейромеревевої моделі

У процесі візуальної інтерпретації було зафіксовано характерну особливість маніпулятивних повідомлень асиметрію емоційного навантаження між початком та завершенням тексту, що свідчить про поступове нарощування психологічного тиску. Побудована візуальна карта таких текстів відображає структурні зони емоційної напруги, які корелюють із типами маніпуляцій експресивною, емпатійною або страхоцентричною.

Розроблений підхід є не лише технічним рішенням, а й методологічним зрушенням у напрямі когнітивної інтерпретації цифрових комунікацій. Його застосування дозволяє перейти від суто статистичного аналізу до виявлення причинно-наслідкових зв'язків між мовними структурами та емоційними ефектами. Це створює підґрунтя для розроблення систем моніторингу інформаційного простору, що спроможні не лише фіксувати факти емоційного впливу, а й пояснювати їх природу, підвищуючи рівень прозорості та довіри до інтелектуальних технологій.

З теоретичного погляду результати дослідження розширюють межі розуміння процесів емоційної семантики у цифрових текстах і сприяють формуванню нової парадигми обробки природної мови, орієнтованої на інтерпретованість і візуальну прозорість. У прикладному вимірі метод може бути використаний для побудови систем виявлення маніпуляцій у медіа-контенті, психологічного аналізу користувацьких повідомлень, оцінки довіри до інформаційних джерел і в освітніх технологіях, де важливо забезпечити пояснюваність роботи ШІ-систем. Такий підхід створює основу для розроблення когнітивно орієнтованих систем цифрової безпеки, які об'єднують інструменти емоційного аналізу, нейромережевої інтерпретації та візуальної аналітики.

#### Список використаних джерел

[1] E. M. Nisotaki and C. Mourlas, “Digital Manipulation: The Role of Computational Propaganda in Influencing the Public Sphere,” in *Digital Populism and the Use of Neo-Propaganda and Fake News*, IGI Global Scientific Publishing, 2025, pp. 255–300. [Online]. Available: <https://www.igi-global.com/chapter/digitalmanipulation/378964> [Accessed: Oct. 22, 2025].

[2] D. Yurchenko, O. Mazurets, V. Didur, and M. Molchanova, “Approach to Using Cloud Services for Visual Analytics of Neural Network Analysis of Texts Emotional Tonality,” in *Proceedings of the XLVII International Scientific and Practical Conference “The Future of Scientific Discoveries: New Trends and Technologies”*, Marseille, France, Nov. 13–15, 2024, pp. 108–113. [Online]. Available: [https://isu-conference.com/wp-content/uploads/2024/11/The\\_future\\_of\\_scientific\\_discoveries\\_new\\_trends\\_and\\_technologies\\_November\\_13\\_15\\_2024\\_Marseille\\_France.pdf](https://isu-conference.com/wp-content/uploads/2024/11/The_future_of_scientific_discoveries_new_trends_and_technologies_November_13_15_2024_Marseille_France.pdf) [Accessed: Oct. 22, 2025].

[3] Д. Ю. Юрченко, О. В. Мазурець, О. О. Залуцька, і Ю. Г. Безпрозвана, “Підхід до візуального пояснення результатів нейромережевого аналізу емоційної тональності повідомлень у соціальних мережах,” *Матеріали XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024»*, Хмельницький, 15–16 листопада 2024 р., с. 565–571. [Online]. Available: <https://kn.khmn.edu.ua/wp-content/uploads/sites/18/apkn-2024-corporpaper.pdf> [Accessed: Oct. 22, 2025].

[4] Д. Ю. Юрченко, О. М. Овчарук, О. В. Мазурець, і П. О. Шевчук, “Метод використання нейромережі гібридної архітектури для визначення емоційної тональності текстових повідомлень” *Вимірвальна та обчислювальна техніка в технологічних процесах*, № 2, с. 136–141, 2025. DOI: 10.31891/2219-9365-2025-82-18.