УДК 004.8

Малярчук Н.В.[1], Молчанова М.О.[2]

[1] Комунальний заклад загальної середньої освіти «Ліцей №17 Хмельницької міської ради»
[2] Хмельницький національний університет

ПІДХІД ДО НЕЙРОМЕРЕЖЕВОГО ВИЯВЛЕННЯ ОЗНАК НАСИЛЬСТВА ГЕНДЕРНОГО СПРЯМУВАННЯ ЗА ПОВІДОМЛЕННЯМИ СОЦІАЛЬНО-ОРІЄНТОВАНИХ СЕРВІСІВ

Робота присвячена нейромережевому виявленню ознак насильства гендерного спрямування за текстами повідомлень соціально-орієнтованих сервісів. Запропоновано підхід, у якому трансформерні моделі глибокого навчання аналізують лінгвістичні та контекстуальні патерни, що відображають вербальну агресію, приниження, погрози та контроль, пов'язані з гендерною нерівністю, з урахуванням багатозначності висловлювань і впливу контексту діалогу. Забезпечується можливість подальшої інтеграції модуля в системи моніторингу та підтримки постраждалих, що сприяє більш ранньому виявленню ризикованих ситуацій і підвищенню ефективності цифрових сервісів соціальної допомоги.

The paper focuses on a neural network–based approach to detecting signs of gender-based violence in messages from social-oriented online services. The proposed method employs transformer-based deep learning models to capture linguistic and contextual patterns that reflect verbal aggression, humiliation, threats, and controlling behaviour linked to gender inequality, while accounting for ambiguity and dialogue context. The resulting module can be integrated into monitoring and support systems for victims, enabling earlier detection of high-risk situations and improving the effectiveness of digital social assistance services.

Текстові звернення до соціально-орієнтованих сервісів, чат-ботів і онлайн-консультацій дедалі частіше містять опис ситуацій, пов'язаних з насильством гендерного спрямування, однак подаються у вигляді коротких, фрагментарних повідомлень [1]. У таких повідомленнях постраждалі нерідко уникають прямої кваліфікації пережитого як насильства, натомість описують окремі епізоди, емоційні стани, висловлювання партнера чи близького оточення [2]. Методи автоматизованого аналізу тексту, що спираються на ключові слова або прості статистичні моделі, виявляються недостатньо чутливими до завуальованих, контекстно насичених форм вербального насильства, приниження, контролю й погроз [3]. Це зумовлює потребу у нейромережевих підходах, здатних виявляти ознаки гендерно обумовленого насильства на рівні одного окремого повідомлення, без залучення повної історії діалогу.

Поширення онлайн-комунікацій та соціально-орієнтованих сервісів значно збільшило обсяг текстового контенту, у якому можуть проявлятися ознаки

насильства гендерного спрямування. Виявлення таких проявів є важливим для створення безпечного цифрового середовища, захисту прав користувачів та запобігання психологічній шкоді [4]. Традиційні підходи до модерації контенту, що покладаються на людську експертизу, часто є неефективними через швидкість поширення повідомлень, що робить актуальним застосування автоматизованих методів аналізу тексту [5].

Сучасні технології обробки природної мови дозволяють ефективно розпізнавати лексичні [6], синтаксичні [7] та семантичні [8] маркери насильницьких або дискримінаційних висловлювань щодо певної гендерної групи. Використання нейромережевих моделей, зокрема трансформерних архітектур [9], забезпечує контекстне розуміння повідомлень [10], виявлення прихованих форм образ, погроз чи психологічного тиску [11], які можуть бути замасковані або виражені непрямими конструкціями. Такі моделі дозволяють аналізувати великі обсяги даних у реальному часі та забезпечують масштабованість при роботі з соціальними платформами [12].

Метою роботи є розроблення нейромережевого підходу до автоматизованого виявлення ознак насильства гендерного спрямування за одним текстовим повідомленням користувача соціально-орієнтованих цифрових сервісів. Об'єктом дослідження є процес обробки індивідуальних текстових звернень у системах онлайн-підтримки, консультацій та сервісах соціальної допомоги. Предметом дослідження виступають моделі та методи нейромережевої класифікації тексту, орієнтовані на розпізнавання лінгвістичних маркерів гендерно обумовленого насильства у форматі одиничного повідомлення.

Запропонований підхід ґрунтується на донавчанні попередньо натренованої трансформерної мовної моделі на спеціалізованому корпусі текстів, у якому повідомлення позначені за наявністю або відсутністю ознак насильства гендерного спрямування, а за потреби – за підтипами проявів (вербальні образи, приниження, погрози, контроль доступу до ресурсів тощо). Текстове повідомлення проходить етапи нормалізації та токенізації, після чого перетворюється на контекстно-залежне векторне представлення, що подається на вхід класифікаційному шару нейромережі. Модель навчається відносити кожне окреме повідомлення до класу «містить ознаки гендерно обумовленого насильства» чи «не містить», спираючись на сукупність лексичних, синтаксичних і семантичних патернів, притаманних описам токсичних і небезпечних взаємодій.

Для підвищення практичної придатності розроблюваного рішення передбачається використання інтерпретованих виходів моделі на рівні важливості окремих слів та словосполучень, що формують підґрунтя для експертної верифікації спрацьовувань. Це дозволяє фахівцям соціальної та психологічної допомоги оцінювати не лише факт спрацювання класифікатора, а й ті текстові фрагменти, які стали вирішальними для віднесення повідомлення до категорії ризику. Очікується, що інтеграція такого нейромережевого модуля у цифрові сервіси соціальної підтримки сприятиме ранньому виявленню потенційно небезпечних ситуацій,

оперативному скеруванню постраждалих до відповідних служб та підвищенню ефективності моніторингу проявів насильства гендерного спрямування у цифровому середовищі.

Перелік посилань

1. García-Rojas, A. D., Gómez, A. H., Montero-Fernández, D., & Rodríguez-Vargas, S. (2024). Perception of University Students Regarding Gender-Based Violence: Identification, Analysis and Detection. Sexes, 5(4), 758-768.
2. Wagner, A., & Condello, A. (Eds.). (2025). (In) Visible Signs of Gender-Based Violence (Vol. 1). Springer Nature.
3. Bondestam, F. (2024). Addressing Gender-Based violence through the ERA policy framework: a systemic solution to dilemmas and contestations for institutions. International Journal of Higher Education, 13(2), 74.
4. 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-орієтовану інформаційну систему. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2024. №1 (331). С. 101-106.
5. Денисенко Б.О., Молчанова М.О., Мазурець О.В. Інтелектуальна система виявлення дезінформації з застосуванням штучних нейронних мереж. Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024». 15-16 листопада 2024. Хмельницький, 2024. с. 167-174.
6. Sobko O., Mazurets O., Didur V., Chervonchuk I. Recurrent Neural Network Model Architecture for Detecting a Tendency to Atypical Behavior Of Individuals by Text Posts. Theoretical and Practical Aspects of Modern Research. Proceedings of XXVI International scientific and practical conference. June 5-7, 2024. International Scientific Unity. Ottawa, Canada. 2024. Pp. 113-117.
7. 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.
8. Yurchenko D., Mazurets O., Didur V., Molchanova M. Approach to Using Cloud Services for Visual Analytics of Neural Network Analysis of Texts Emotional Tonality. The Future of Scientific Discoveries: New Trends and Technologies. Proceedings of the XLVII International scientific and practical conference. November 13-15, 2024. Marseille, France. 2024. Pp. 108-113.
9. 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.
10. O. Mazurets, R. Vit, M. Molchanova, I. Tymofiiev, O. Sobko, Context-enriched approach to students depression monitoring in education using BERT-GPT hybrid model, CEUR Workshop Proceedings 4096 (2025) 167-176.
11. Molchanova M., Didur V., Sobko O., Mazurets O. Detection of Web Propaganda Patterns by Transformer Neural Networks: Improving Efficiency via Dataset Balancing, CEUR Workshop Proceedings, 2025, vol. 3988, pp. 112-126.
12. E. A. Manziuk, O. V. Sobko, I. O. Podhorniuk, M. O. Molchanova, O. V. Mazurets, Multifactorial analysis of mobbing behavioral signs in educational environments posts by NLP means, Journal of Physics Conference Series 3105(1) (2025) 012025.