

ХМЕЛЬНИЦЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
МІНІСТЕРСТВА ОСВІТИ І НАУКИ УКРАЇНИ

Кваліфікаційна наукова
праця на правах рукопису

ОВЧАРУК ОЛЕКСАНДР МИКОЛАЙОВИЧ

УДК 004.8

ДИСЕРТАЦІЯ

МЕТОДИ АНАЛІЗУ ПСИХІЧНОГО СТАНУ В ГРОМАДЯН ІЗ
ПОСТТРАВМАТИЧНИМ СТРЕСОВИМ РОЗЛАДОМ ЗА
ІНТЕЛЕКТУАЛЬНИМ АНАЛІЗОМ КОРИСТУВАЦЬКОГО КОНТЕНТУ

122 Комп'ютерні науки
(шифр і назва спеціальності)

12 Інформаційні технології
(галузь знань)

Подається на здобуття наукового ступеня доктора філософії

Дисертація містить результати власних досліджень. Подані до захисту наукові положення є власним напрацюванням. Всі використані ідеї, наукові результати, цитати супроводжуються належними посиланнями на їх авторів та джерела опублікування. Всі частини тексту дисертації, під час написання яких використовувались технології штучного інтелекту, перевірені та відредаговані особисто.


(підпис) Овчарук Олександр Миколайович

Науковий керівник:

Мазурець Олександр Вікторович, кандидат технічних наук, доцент

АНОТАЦІЯ

Овчарук О. М. Методи аналізу психічного стану в громадян із посттравматичним стресовим розладом за інтелектуальним аналізом користувацького контенту. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії з галузі знань 12 Інформаційні технології за спеціальністю 122 Комп'ютерні науки. Хмельницький національний університет, Хмельницький, 2026.

Актуальність теми дослідження. У сучасному світі стрімке зростання соціальних мереж і цифрового контенту створює нові можливості для вивчення психічного стану людини без прямого втручання в її особистий простір. Особливо це важливо в умовах зростання кількості осіб, які страждають від посттравматичного стресового розладу, зокрема серед військовослужбовців, ветеранів, волонтерів та цивільного населення, постраждалого внаслідок збройних конфліктів.

ПТСР є серйозною проблемою громадського здоров'я, яка значно впливає на соціальну адаптацію, емоційний стан і якість життя людини. При цьому не всі постраждалі звертаються по медичну допомогу, що ускладнює діагностику та своєчасне втручання.

У зв'язку з цим інтелектуальний аналіз користувацького контенту – таких як пости, коментарі, відео та повідомлення в соціальних мережах – стає перспективним методом раннього виявлення ознак психічних розладів. Використання алгоритмів штучного інтелекту та машинного навчання дозволяє проводити аналіз великих обсягів даних, виявляючи мовні та поведінкові патерни, характерні для ПТСР.

Таким чином, поєднання цифрових технологій і методів скринінгу психічного стану відкриває нові горизонти для створення інструментів моніторингу психічного здоров'я, що є надзвичайно актуальним у період соціальних потрясінь та війни.

Об'єкт дослідження: процес виявлення ознак посттравматичного стресового розладу, супутніх психічних розладів та ризиків соціально небезпечної поведінки громадян за аналізом користувацького текстового контенту.

Предмет дослідження: методи та засоби інтелектуального аналізу користувацького контенту, обробки природної мови та нейромережевого моделювання для виявлення посттравматичного стресового розладу, психічних розладів і поведінкових ризиків.

Методи дослідження. Використано методи обробки природної мови, інтелектуального аналізу даних, машинного та глибокого навчання, трансформерні архітектури, методи багатоміткової класифікації, методи статистичного аналізу й оцінювання якості класифікаційних моделей, а також засоби візуальної інтерпретації результатів нейромережевого аналізу.

Метою дослідження є підвищення точності виявлення ознак посттравматичного стресового розладу і супутніх психічних розладів та розширення можливостей оцінювання ризиків соціально небезпечної поведінки громадян за інтелектуальним аналізом користувацького контенту засобами контекстно-орієнтованих нейромережевих моделей і розширення простору ознак.

Під точністю виявлення розуміються показники класифікації на рівні повідомлення: Accuracy, Precision, Recall, F1-score. Розширення можливостей оцінювання ризиків полягає в переході від рішень щодо окремих повідомлень до скринінгового часового індексу ризику на рівні користувача, який інтегрує оцінки прояву ПТСР, інших психічних розладів та емоційного контексту повідомлень. Супутніми названо психічні розлади, ознаки яких можуть виявлятися в текстах одночасно з ознаками ПТСР; їх зв'язок саме з ПТСР на використаних наборах даних не перевірявся.

Наукова новизна одержаних результатів:

1) удосконалено метод аналізу прояву посттравматичного стресового розладу в користувацькому контенті, який відрізняється від існуючих формуванням спеціалізованого навчального набору даних, нецільовий клас якого

містить тексти тематично суміжних спільнот, зокрема з проявами інших психічних розладів, у поєднанні з контекстно-позиційним представленням тексту, що дало змогу підвищити точність виявлення ознак ПТСР.

2) удосконалено метод виявлення психічних розладів за аналізом користувацького контенту, який відрізняється від існуючих декомпозицією задачі на множину незалежних бінарних класифікаторів з окремо сформованими навчальними вибірками та індивідуальними порогоми прийняття рішення, що не потребує вибору одного домінуючого розладу і не знижує точність порівняно з багатокласовою класифікацією.

3) вперше розроблено метод розширення множини ознак користувацького контенту для попередження соціально небезпечної поведінки громадян із посттравматичним стресовим розладом, який відрізняється від існуючих підходів інтеграцією оцінок прояву ПТСР, інших психічних розладів та емоційного контексту повідомлень у пояснюваний часовий індекс ризику на рівні користувача, що дає змогу формувати скринінгову оцінку ризику соціально небезпечної поведінки.

Основний зміст роботи. У першому розділі проаналізовано сучасні методи, засоби та технології автоматизованого виявлення ПТСР та інших психічних розладів за користувацьким текстовим контентом, ознаки психічних розладів і соціально небезпечної поведінки в текстах, а також підходи до пояснення нейромережових рішень. Встановлено, що відомі рішення не передбачають контрольованого відмежування ознак ПТСР від тематично суміжних станів на рівні дописів соціальних мереж, орієнтовані переважно на виявлення одного домінуючого розладу та оцінюють поведінкові ризики без урахування їх часової динаміки. Сформульовано мету та завдання дослідження.

У другому розділі формалізовано предметну область і загальну постановку задачі, подано теоретичні основи трьох запропонованих методів: методу аналізу прояву ПТСР у користувацькому контенті, методу виявлення психічних розладів за аналізом користувацького контенту та методу

розширення множини ознак користувацького контенту для попередження соціально небезпечної поведінки громадян із ПТСР, а також показано їх взаємозв'язок у комплексному аналізі.

У третьому розділі наведено алгоритмічну реалізацію запропонованих методів: нейромережеву модель на основі DeBERTa для виявлення ПТСР та окрему експериментальну модель із квантовим шаром; множину незалежних бінарних моделей DistilBERT з індивідуальними порогами за індексом Юдена для багатоміткового виявлення психічних розладів; розширений ознаковий профіль і часовий індекс ризику. Описано формування спеціалізованих наборів даних, метрики оцінювання, дизайн експериментальних досліджень та етичні аспекти застосування автоматизованих рішень.

У четвертому розділі наведено результати експериментального дослідження. Для методу аналізу прояву ПТСР отримано Accuracy 0,934, Precision 0,948 та F1-score 0,933 на валідаційній вибірці; в окремому пошуковому експерименті досліджено модель із квантовим шаром (Accuracy 0,85). Для методу виявлення психічних розладів найкраща конфігурація з адаптивним порогоуванням забезпечила Macro Accuracy 0,9316, Macro Precision 0,9338, Macro Recall 0,9296 та Macro F1-score 0,9296, а Accuracy окремих бінарних моделей становила від 0,87 до 1,00. Для методу розширення множини ознак реалізовано розширений ознаковий профіль і часовий індекс ризику, який інтегрує оцінки ПТСР, супутніх розладів та емоційного контексту повідомлень. Пояснюваність нейромережевих рішень досліджено засобами Transformers Interpret і BertViz; окреслено обмеження запропонованих методів і використаних наборів даних.

Практичне значення отриманих результатів полягає у програмній реалізації комплексного підходу до інтелектуального аналізу користувацького контенту з метою виявлення ознак посттравматичного стресового розладу та скринінгового оцінювання ризику соціально небезпечної поведінки. Розроблена система аналізу дозволяє автоматизовано обробляти великі обсяги текстової

інформації, виявляючи мовні, емоційні та поведінкові патерни, характерні для ПТСР. Комплексний підхід включає багаторівневу оцінку ризиків із числовими показниками інтенсивності виявлених ознак та ранжуванням потенційно ризикових поведінкових патернів.

Впровадження такої системи може бути ефективним інструментом у практиці психологічного моніторингу, допомоги фахівцям та соціальним службам у вчасному виявленні осіб, які потребують підтримки.

Результати дисертаційної роботи були впроваджені в освітній процес Хмельницького національного університету у процесі викладання дисципліни «Моделі та методи текстової аналітики». Апробація та впровадження розробок здійснені в діяльності ГО «ІТ Кластер м. Хмельницького», ПП «Авіві» та ТОВ «Стфалкон», а також при виконанні держбюджетних тем Хмельницького національного університету: «Розробка методів та алгоритмів ідентифікації та класифікації залишків зруйнованих будівель на основі нейромережевого аналізу візуальних даних» (ДР № 0125U003585) та «Нейромережеві методи класифікації текстильних відходів для систем циркулярної економіки» (ДР № 0126U002064).

Зокрема, у межах держбюджетної теми «Розробка методів та алгоритмів ідентифікації та класифікації залишків зруйнованих будівель на основі нейромережевого аналізу візуальних даних» (ДР № 0125U003585) результати дисертації впроваджено на рівні методичного та алгоритмічного забезпечення побудови інтелектуальних систем класифікації: використано підходи до формування спеціалізованих навчальних наборів даних, адаптації нейромережевих моделей до особливостей предметної області, побудови багаторівневих схем класифікації та оцінювання якості моделей за сукупністю метрик.

У межах держбюджетної теми «Нейромережеві методи класифікації текстильних відходів для систем циркулярної економіки» (ДР № 0126U002064) результати дисертаційної роботи впроваджено в частині розроблення моделей багатокласової класифікації, розширення множини

інформативних ознак об'єктів аналізу, а також програмної реалізації модулів оцінювання та візуалізації результатів класифікації.

Особистий внесок здобувача. Усі наукові результати, викладені в дисертаційній роботі, отримані здобувачем особисто. У роботах, опублікованих у співавторстві, здобувачеві належать: метод аналізу прояву посттравматичного стресового розладу в користувацькому контенті на основі контекстно-позиційного представлення тексту; принцип формування спеціалізованого помилко-орієнтованого навчального набору даних, нецільовий клас якого містить тексти з проявами інших психічних розладів; метод виявлення психічних розладів за аналізом користувацького контенту на основі роздільного підходу до класифікації з можливістю ідентифікації кількох розладів одночасно; метод розширення множини ознак користувацького контенту та підхід до скринінгового оцінювання ризику соціально небезпечної поведінки громадян із ПТСР; програмна реалізація та експериментальні дослідження розроблених методів і засобів.

Ключові слова: посттравматичний стресовий розлад, ПТСР, інтелектуальний аналіз користувацького контенту, аналіз психічного стану, обробка природної мови, нейронні мережі, трансформери, DeBERTa, багатоміткова класифікація, соціально небезпечна поведінка, пояснюваність.

ANNOTATION

Ovcharuk O. M. Methods of analyzing the mental state of citizens with post-traumatic stress disorder based on intelligent analysis of user content. – Qualifying scientific manuscript.

Thesis for the degree of Doctor of Philosophy in the field of knowledge 12 Information Technologies, specialty 122 Computer Science. – Khmelnytskyi National University, Khmelnytskyi, 2026.

Relevance of the research topic. In the modern world, the rapid growth of social networks and digital content creates new opportunities for studying the

mental state of a person without direct interference in their personal space. This is especially important in the context of the growing number of people suffering from post-traumatic stress disorder, in particular among military personnel, veterans, volunteers and civilians affected by armed conflicts.

PTSD is a serious public health problem that significantly affects social adaptation, emotional state and quality of life of a person. At the same time, not all victims seek medical help, which complicates diagnosis and timely intervention.

In this regard, the intelligent analysis of user content – such as posts, comments, videos and messages on social networks – is becoming a promising method for early detection of signs of mental disorders. The use of artificial intelligence and machine learning algorithms allows for the analysis of large volumes of data, revealing speech and behavioral patterns characteristic of PTSD.

Thus, the combination of digital technologies and mental state screening methods opens up new horizons for the creation of mental health monitoring tools, which is extremely relevant in times of social upheaval and war.

Object of research: the process of detecting signs of post-traumatic stress disorder, comorbid mental disorders, and risks of socially dangerous behavior in citizens through the analysis of user-generated textual content.

Subject of research: methods and tools of intelligent user content analysis, natural language processing, and neural network modeling for detecting post-traumatic stress disorder, mental disorders, and behavioral risks.

Research methods. Methods of natural language processing, data mining, machine and deep learning, transformer architectures, multi-label classification methods, methods of statistical analysis and quality evaluation of classification models, as well as tools for visual interpretation of the results of neural network analysis were used.

The purpose of the research is to improve the accuracy of detecting signs of post-traumatic stress disorder and comorbid mental disorders and to extend the capabilities of assessing the risks of socially dangerous behavior of citizens

through the intelligent analysis of user content by means of context-oriented neural network models and feature space expansion.

Detection accuracy is understood as message-level classification metrics: Accuracy, Precision, Recall and F1-score. Extending the capabilities of risk assessment means moving from decisions on individual messages to a user-level temporal screening risk index that integrates the assessments of PTSD manifestation, other mental disorders and the emotional context of messages. Comorbid disorders are mental disorders whose signs may appear in texts together with signs of PTSD; their association with PTSD specifically was not verified on the datasets used.

Scientific novelty of the obtained results:

1) the method for analyzing the manifestation of post-traumatic stress disorder in user content has been improved, which differs from the existing ones by forming a specialized training dataset whose non-target class contains texts from thematically related communities, including those with manifestations of other mental disorders, in combination with a contextual-positional text representation, which made it possible to improve the accuracy of detecting signs of PTSD.

2) the method for detecting mental disorders by analyzing user content has been improved, which differs from the existing ones by decomposing the task into a set of independent binary classifiers with separately formed training samples and individual decision thresholds, which does not require choosing a single dominant disorder and does not reduce accuracy compared to multi-class classification.

3) a method for expanding the set of user content features to prevent socially dangerous behavior of citizens with post-traumatic stress disorder has been developed for the first time, which differs from existing approaches by integrating the assessments of PTSD manifestation, other mental disorders and the emotional context of messages into an explainable user-level temporal risk index, which makes it possible to form a screening assessment of the risk of socially dangerous behavior.

Main content of the work. The first chapter analyzes modern methods, tools and technologies for automated detection of PTSD and other mental disorders in

user-generated textual content, the features of mental disorders and socially dangerous behavior in texts, as well as approaches to explaining neural network decisions. It was established that known solutions do not provide controlled separation of PTSD signs from thematically related conditions at the level of social media posts, are mostly focused on detecting a single dominant disorder, and assess behavioral risks without considering their temporal dynamics. The purpose and objectives of the research are formulated.

The second chapter formalizes the subject area and the general problem statement and presents the theoretical foundations of the three proposed methods: the method for analyzing the manifestation of PTSD in user content, the method for detecting mental disorders by analyzing user content, and the method for expanding the set of user content features to prevent socially dangerous behavior of citizens with PTSD; their interrelation within the comprehensive analysis is shown.

The third chapter presents the algorithmic implementation of the proposed methods: a DeBERTa-based neural network model for PTSD detection and a separate experimental model with a quantum layer; a set of independent binary DistilBERT models with individual thresholds based on the Youden index for multi-label detection of mental disorders; an extended feature profile and a temporal risk index. The formation of specialized datasets, evaluation metrics, the design of experimental studies and the ethical aspects of applying automated decisions are described.

The fourth chapter presents the results of the experimental study. For the method of analyzing the manifestation of PTSD, Accuracy of 0.934, Precision of 0.948 and F1-score of 0.933 were obtained on the validation set; a model with a quantum layer (Accuracy of 0.85) was studied in a separate exploratory experiment. For the method of detecting mental disorders, the best configuration with adaptive thresholding provided Macro Accuracy of 0.9316, Macro Precision of 0.9338, Macro Recall of 0.9296 and Macro F1-score of 0.9296, while the Accuracy of individual binary models ranged from 0.87 to 1.00. For the method of

expanding the set of features, an extended feature profile and a temporal risk index that integrates the assessments of PTSD, comorbid disorders and the emotional context of messages were implemented. The explainability of neural network decisions was studied using Transformers Interpret and BertViz; the limitations of the proposed methods and the datasets used are outlined.

The practical significance of the obtained results lies in the programmatic implementation of a comprehensive approach to the intelligent analysis of user content in order to identify signs of post-traumatic stress disorder and for screening assessment of the risk of socially dangerous behavior. The developed analysis system allows for automated processing of large volumes of text information, identifying linguistic, emotional and behavioral patterns characteristic of PTSD. The comprehensive approach includes a multi-level risk assessment with numerical indicators of the intensity of the detected signs and ranking of potentially risky behavioral patterns.

The implementation of such a system can be an effective tool in the practice of psychological monitoring, helping specialists and social services in the timely identification of individuals in need of support.

The results of the dissertation work were implemented in the educational process of Khmelnytskyi National University when teaching the discipline "Models and Methods of Text Analytics". The development was tested and implemented in the activities of the NGO "IT Cluster of Khmelnytskyi", PE "Avivi" and LLC "Stfalcon", as well as in the implementation of the state budget topics of Khmelnytskyi National University: "Development of methods and algorithms for identification and classification of remains of destroyed buildings based on neural network analysis of visual data" (DR No. 0125U003585) and "Neural network methods for classification of textile waste for circular economy systems" (DR No. 0126U002064).

In particular, within the framework of the state budget topic "Development of methods and algorithms for identification and classification of remains of destroyed buildings based on neural network analysis of visual data" (DR No.

0125U003585), the results of the dissertation were implemented at the level of methodological and algorithmic support for building intelligent classification systems. The approaches to forming specialized training datasets, adapting neural network models to domain-specific features, building multi-level classification schemes and evaluating model quality by a set of metrics were used.

Within the framework of the implementation of the state budget topic "Neural network methods for textile waste classification for circular economy systems" (DR No. 0126U002064), the results of the dissertation work were implemented in terms of developing multi-class classification models, expanding the set of informative features of analysis objects, as well as software implementation of modules for evaluating and visualizing classification results.

Personal contribution of the applicant. All scientific results presented in the dissertation were obtained by the applicant personally. In the works published in co-authorship, the applicant owns: the method for analyzing the manifestation of post-traumatic stress disorder in user content based on the contextual-positional representation of the text; the principle of forming a specialized error-oriented training dataset, the non-target class of which contains texts with manifestations of other mental disorders; the method for detecting mental disorders by analyzing user content based on a separate approach to classification with the ability to identify several disorders at once; the method for expanding the set of user content features and the approach to screening assessment of the risk of socially dangerous behavior of citizens with PTSD; software implementation and experimental studies of the developed methods and tools.

Keywords: post-traumatic stress disorder, PTSD, intelligent analysis of user content, mental state analysis, natural language processing, neural networks, transformers, DeBERTa, multi-label classification, socially dangerous behavior, explainability.

СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА ЗА ТЕМОЮ ДИСЕРТАЦІЇ

Статті у наукових виданнях, включених до Переліку наукових фахових видань України:

1. Овчарук О.М., Мазурець О.В. Нейромережеве діагностування проявів ПТСР у текстовому контенті з використанням помилко-орієнтованого навчального набору даних. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2024. №6, Т.1 (343). С. 195-200. <http://doi.org/10.31891/2307-5732-2024-343-6-30>.
2. Овчарук О.М., Мазурець О.В. Нейромережевий метод діагностування психологічних розладів за аналізом повідомлень на основі роздільного підходу до класифікації. Міжнародний науково-технічний журнал «Вимірювальна та обчислювальна техніка в технологічних процесах». 2025. № 1. С. 210-216. <https://doi.org/10.31891/2219-9365-2025-81-25>.
3. Овчарук О.М., Мазурець О.В. Превенція деструктивної поведінки за текстами користувачів з ПТСР методами обробки природної мови. Вісник Хмельницького національного університету. Серія: Технічні науки. 2025. №6, Т.2 (359). С. 458–462. <https://doi.org/10.31891/2307-5732-2025-359-136>.

Праці, які засвідчують апробацію матеріалів дисертації:

4. Krak I., Ovcharuk O., Mazurets O., Molchanova M., Sobko O., Tyschenko O., Barmak O. Method for post-traumatic stress disorder manifestation analyzing in text content. CEUR Workshop Proceedings. 2025. Vol. 3976. P. 206-218. URL: <https://ceur-ws.org/Vol-3976/paper15.pdf>
5. Krak I., Mazurets O., Ovcharuk O., Molchanova M., Barmak O., Azarova L. Transformer-based multilabel classification for identifying hidden psychological conditions in online posts. CEUR Workshop Proceedings. 2025. Vol. 4004. P. 86-97. URL: <https://ceur-ws.org/Vol-4004/paper8.pdf>
6. Krak I., Ovcharuk O., Mazurets O., Sobko O., Molchanova M., Barmak O. Neural Network Assessment of Post-Traumatic Risk Behavior Using Social

Media Posts Analysis. CEUR Workshop Proceedings. 2025. Vol. 4164. P. 141-151. URL: <https://ceur-ws.org/Vol-4164/paper9.pdf>

7. Овчарук О.М., Мазурець О.В. Метод аналізу проявів посттравматичного стресового розладу за нейромережевим аналізом текстів. Матеріали XIII Всеукраїнської науково-практичної конференції «Глушковські читання». 6 грудня 2024. Київ, 2024. С. 218-221.

8. Мазурець О.В., Овчарук О.М. Діагностування проявів посттравматичного стресового розладу за нейромережевим аналізом текстового контенту. Збірник наукових праць XXIII Міжнародної наукової конференції «Нейромережні технології та їх застосування НМТІЗ-2024». 11-12 грудня 2024. Краматорськ-Тернопіль, ДДМА. 2024. С. 85-89.

9. Овчарук О.М., Мазурець О.В. Підхід до оцінювання поведінкових ризиків користувачів з ПТСР за цифровим профілем засобами NLP. Тези доповідей XXIII Міжнародної науково-практичної конференції «Математичне та програмне забезпечення інтелектуальних систем МПЗІС-2025». 19-21 листопада 2025. Дніпро: ДНУ, 2025. С. 241-242.

Публікації, які додатково відображають наукові результати дисертації:

10. Мазурець О.В., Овчарук О.М. Комп'ютерна програма «Інтелектуальна інформаційна система для аналізу прояву посттравматичного стресового розладу в текстовій інформації»: свідоцтво про реєстрацію авторського права № 135968 від 08.05.2025.

11. Мазурець О.В., Овчарук О.М. Комп'ютерна програма «Інтелектуальна інформаційна система для виявлення психічних розладів за аналізом текстової інформації»: свідоцтво про реєстрацію авторського права № 135967 від 08.05.2025.

12. Mazurets O.V., Ovcharuk O.M. Methodology for Identifying Post-Traumatic Stress Disorder Indicators in Text Data. Information Control Systems and Intelligent Technologies. Advances and Applications. Monograph. Liha-Pres. 2025. P. 43-58. ISBN 978-966-397-538-2. <https://doi.org/10.36059/978-966-397-538-2-3>.

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ.....	18
ВСТУП	20
РОЗДІЛ 1. АНАЛІЗ МЕТОДІВ, ЗАСОБІВ ТА ТЕХНОЛОГІЙ ДЛЯ АВТОМАТИЗОВАНОГО ВИЯВЛЕННЯ ПТСР У КОРИСТУВАЦЬКОМУ ТЕКСТОВОМУ КОНТЕНТІ	30
1.1. Сучасний стан інтелектуального аналізу текстового контенту для автоматизованого виявлення ПТСР	30
1.2. Сучасний стан виявлення психічних розладів за текстовим контентом.....	34
1.3. Урядові програми та нормативні документи	40
1.4. Аналіз ознак психічних розладів у текстовому контенті	41
1.5. Аналіз ознак користувацького контенту для попередження соціально небезпечної поведінки	45
1.6. Пояснюваність результатів	47
1.7. Мета та завдання дослідження	52
РОЗДІЛ 2. ТЕОРЕТИЧНІ ОСНОВИ ТА ФОРМАЛІЗАЦІЯ МЕТОДІВ АНАЛІЗУ ПСИХІЧНОГО СТАНУ ЗА КОРИСТУВАЦЬКИМ КОНТЕНТОМ .	54
2.1. Формалізація предметної області та загальної постановки задачі	54
2.2. Метод аналізу прояву ПТСР у користувацькому контенті	57
2.3. Метод виявлення психічних розладів за аналізом користувацького контенту	59
2.4. Метод розширення множини ознак користувацького контенту для попередження соціально небезпечної поведінки громадян із посттравматичним стресовим розладом	66
2.5. Взаємозв'язок методів у комплексному аналізі	71
2.6. Висновки до розділу 2	76

РОЗДІЛ 3. АЛГОРИТМІЧНА РЕАЛІЗАЦІЯ ТА МЕТОДИКА ЕКСПЕРИМЕНТАЛЬНОГО ДОСЛІДЖЕННЯ.....	78
3.1. Алгоритмічна реалізація методу аналізу прояву ПТСР у користувацькому контенті	78
3.2. Нейромережеві моделі для виявлення ПТСР.....	79
3.2.1. Нейромережева модель DeBERTa.....	79
3.2.2. Нейромережева модель з квантовим шаром	82
3.3. Алгоритмічна реалізація методу виявлення психічних розладів за аналізом користувацького контенту	87
3.4. Алгоритмічна реалізація методу розширення множини ознак користувацького контенту	92
3.5. Набори даних та їх підготовка.....	96
3.5.1. Дані для виявлення ПТСР	96
3.5.2. Дані для виявлення психічних розладів	99
3.5.3. Дані для часової оцінки ризикової поведінки.....	101
3.6. Метрики оцінювання та функції втрат	101
3.7. Дизайн експериментальних досліджень.....	105
3.8. Етичні аспекти, пояснюваність та межі автоматизованих рішень	107
3.9. Висновки до розділу 3	108
РОЗДІЛ 4. ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ ТА АНАЛІЗ РЕЗУЛЬТАТІВ.....	110
4.1. Програмне середовище експериментів.....	110
4.2. Дослідження методу аналізу прояву ПТСР у користувацькому контенті.	111
4.2.1. Дослідження нейромережевої моделі з квантовим шаром	111
4.2.2. Дослідження нейромережевої моделі DeBERTa	113
4.3. Дослідження методу виявлення психічних розладів за аналізом користувацького контенту	119
4.3.1. Дослідження методу виявлення психічних розладів за аналізом користувацького контенту без адаптивного порогу.....	119

4.3.2. Дослідження методу виявлення психічних розладів за аналізом користувацького контенту з адаптивним порогом	125
4.4. Дослідження методу розширення множини ознак користувацького контенту для попередження соціально небезпечної поведінки	134
4.5. Узагальнення результатів експериментального дослідження	140
4.6. Обмеження запропонованих методів та використаних наборів даних	140
4.7. Висновки до розділу 4	148
ВИСНОВКИ.....	151
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	155
ДОДАТОК А. СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА	179
ДОДАТОК Б. АКТИ ВПРОВАДЖЕННЯ.....	181
ДОДАТОК В. ПРОГРАМНІ КОДИ	185
ДОДАТОК Г. АВТОРСЬКІ СВДОЦТВА.....	187

ПЕРЕЛІК СКОРОЧЕНЬ

- ПТСР – посттравматичний стресовий розлад
- ІІ – штучний інтелект
- МН – машинне навчання
- ML (Machine Learning) – машинне навчання
- DL (Deep Learning) – глибоке навчання
- NLP (Natural Language Processing) – обробка природної мови
- BERT (Bidirectional Encoder Representations from Transformers) – двонапрямні представлення енкодера з трансформерів
- DeBERTa (Decoding-enhanced BERT with disentangled attention) – модифікована архітектура BERT із розділеним механізмом уваги
- CNN (Convolutional Neural Network) – згорткова нейронна мережа
- LSTM (Long Short-Term Memory) – мережа довгої короткочасної пам'яті
- BOW (Bag of Words) – модель «мішок слів»
- TF-IDF (Term Frequency – Inverse Document Frequency) – статистична міра ваги терміна в тексті
- SVM (Support Vector Machine) – метод опорних векторів
- AUC (Area Under the Curve) – площа під кривою
- ROC (Receiver Operating Characteristic) – ROC-крива
- TP (True Positive) – кількість правильно передбачених позитивних випадків
- TN (True Negative) – кількість правильно передбачених негативних випадків
- FP (False Positive) – кількість хибно передбачених позитивних випадків
- FN (False Negative) – кількість хибно передбачених негативних випадків
- F1-score – F1-міра, гармонійне середнє Precision і Recall
- DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition) – діагностичний і статистичний посібник із психічних розладів, 5-те видання
- CAPS-5 (Clinician-Administered PTSD Scale for DSM-5) – шкала клінічного оцінювання ПТСР для DSM-5

AVEC-19 (Audio/Visual Emotion Challenge, 2019) – корпус/змагання з аудіовізуального аналізу емоцій

CB-PTSD (Childbirth-Related/Postpartum PTSD) – посттравматичний стресовий розлад, пов'язаний із пологами

ADA (text-embedding-ada-002) – модель векторних подань тексту

RTD (Replaced Token Detection) – виявлення заміненних токенів

MCML (Multi-Class Multi-Level) – багатокласова багаторівнева класифікація

MCSL (Multi-Class Single-Level) – багатокласова однорівнева класифікація

MLC (Multi-Label Classification) – багатоміткова класифікація

SLC (Single-Label Classification) – класифікація з однією міткою

BR (Binary Relevance) – метод бінарної релевантності

CC (Classifier Chain) – ланцюг класифікаторів

LP (Label Powerset) – метод набору потужності міток

XAI (Explainable Artificial Intelligence) – пояснюваний штучний інтелект

LIME (Local Interpretable Model-Agnostic Explanations) – локальні інтерпретовані модельно-незалежні пояснення

SHAP (SHapley Additive exPlanations) – адитивні пояснення на основі значень Шеплі

Grad-CAM (Gradient-weighted Class Activation Mapping) – карти активації класів, зважені за градієнтом

NRC Emotion Lexicon – лексикон емоцій Національної дослідницької ради Канади

ВСТУП

Актуальність теми. У світлі глибоких соціальних трансформацій, спричинених збройними конфліктами, масштабними кризами та ескалацією стресогенних чинників у суспільстві, проблема виявлення та підтримки осіб із посттравматичним стресовим розладом набуває особливої гостроти. Ураження психіки внаслідок травматичних подій має не лише індивідуальні, але й серйозні суспільні наслідки, зокрема в контексті зростання ризиків соціально небезпечної поведінки, маргіналізації та психоемоційної нестабільності в громадах. За оцінками фахівців, психологічної підтримки потребуватиме від 40 до 50 % населення України. Особливо гостро ця потреба проявляється серед окремих соціальних груп: орієнтовно 1,8 мільйона осіб серед військових та ветеранів, близько 7 мільйонів – серед людей похилого віку, а також майже 4 мільйони – серед дітей і підлітків. Очікується, що загальна кількість звернень із питань психічного здоров'я на рівні первинної медичної допомоги може сягнути 27 мільйонів. При цьому, за прогнозами, приблизно 3-4 мільйони громадян матимуть психічні розлади середньої або тяжкої форми.

Розробка новітніх технологій на перетині штучного інтелекту, цифрової лінгвістики та психології відкриває перспективи для створення інструментів раннього виявлення ознак ПТСР шляхом аналізу користувацького контенту в цифровому середовищі. У цьому контексті інтелектуальні системи набувають статусу важливого елементу підтримки психічного здоров'я населення, що на пряму корелює з досягненням Цілі 3 «Забезпечення здорового способу життя і сприяння добробуту для всіх у будь-якому віці» та Цілі 16 «Сприяння побудові мирного та інклюзивного суспільства» у межах Цілей сталого розвитку ООН.

Таким чином, інноваційні підходи до аналізу цифрової поведінки користувачів, зокрема у сфері виявлення ризиків ПТСР, не лише підвищують

ефективність психосоціальної допомоги, але й сприяють формуванню стійких, безпечних і здорових спільнот.

Виявлення ознак психічного стану, зокрема симптомів посттравматичного стресового розладу, належить до задач обробки природної мови, що набуває все більшої популярності серед українських та зарубіжних дослідників у контексті застосування штучного інтелекту для аналізу цифрового користувацького контенту.

Проблему ПТСР та психічного здоров'я населення України досліджували українські науковці. Клініко-психологічні аспекти ПТСР досліджували І. І. Приходько, який дослідив психологічну реабілітацію військовослужбовців, що брали участь у бойових діях [1]; І. Пінчук, А. Солонський, Ю. Ячник, О. Копчак та співавтори, які проаналізували психологічне благополуччя українських студентів в умовах повномасштабної війни [2]; О. Лушак, М. Великодна, С. Болман та співавтори, які оцінили поширеність стресу, тривоги та симптомів посттравматичного стресового розладу серед українців після першого року російського вторгнення [3]; Д. Марценковський та співавтори, що дослідили стан психічного здоров'я в Україні у 2023 році [4]; І. Кліменко, Н. Твердохліб, О. Злобін та співавтори, які узагальнили підходи до діагностики, лікування та підтримки військовослужбовців із ПТСР [5]; С. Зубченко, І. Кріль, О. Надіжко та співавтори, які висвітлили клінічні й лабораторні зміни при ПТСР [6]. Методи інтелектуального аналізу природномовних текстів, дотичні до задачі дисертації, розвивають О. О. Марченко зі співавторами, які дослідили налаштування BERT-подібних моделей для оцінювання тональності текстів і сформували українськомовний навчальний корпус для аналізу емоційного забарвлення [7, 8]; Х. В. Ліпяніна-Гончаренко зі співавторами, які запропонували інтелектуальний метод класифікації користувацьких відгуків за їх емоційним забарвленням [9]; В. А. Висоцька зі співавторами, які розробили інформаційну технологію класифікації текстів для виявлення ознак ПТСР на основі глибинного навчання [10]; Ю. І. Грицюк і Р. Я.

Корбило, які запропонували систему оцінювання психологічного стану військовослужбовців методами оброблення природної мови [11]; Е. А. Манзюк зі співавторами, які запропонували метод класифікації текстового контенту в багатовимірному просторі ознак та поєднання аналізу тональності з доменними знаннями для медичних текстів [12, 13]; Р. О. Багрій зі співавторами, які розробили метод семантичного аналізу навчальних текстів і метод виявлення образливого мовлення в українськомовних текстах [14, 130]; П. М. Радюк зі співавторами, які розвинули методи пояснюваного глибокого навчання на основі матриць переходу та подання рішень моделей у зрозумілих користувачеві ознаках [15, 16]; І. В. Крак, О. В. Бармак зі співавторами, які дослідили нейромережеве виявлення психічних станів та образливого мовлення в текстовому контенті [130, 158, 159]. Водночас серед іноземних дослідників вагомий внесок у розроблення методів автоматизованого виявлення ПТСР і психічних розладів за текстовим та користувацьким контентом зробили J. Sawalha [17], R. Quillivic [18], A. Bartal [19], D. Owen [20], M. Malgaroli [21], G. Gkotsis [22], J. Kim [23] та інші, праці яких присвячені застосуванню методів машинного навчання, глибокого навчання та обробки природної мови для аналізу психічного стану людини за мовними маркерами, інтерв'ю та контентом соціальних мереж.

Аналіз існуючих методів виявлення психічних розладів, зокрема посттравматичного стресового розладу, за користувацьким контентом у цифровому середовищі показав їхню здатність до виявлення загальних ознак емоційної нестабільності, тривожності чи депресії. Проте ці методи здебільшого не враховують контекстно залежні прояви ПТСР, що ускладнює відмежування цього стану від інших психічних порушень із подібною симптоматикою. Крім того, більшість існуючих моделей орієнтовані на виявлення лише одного домінуючого розладу, що не відповідає реаліям, у яких часто спостерігається коморбідність (одночасна наявність кількох розладів).

Методи, засновані переважно на аналізі лінгвістичних характеристик, не забезпечують комплексного урахування емоційно-семантичних і поведінкових патернів, які мають ключове значення у виявленні ризиків соціально небезпечної поведінки, зумовленої ПТСР. Також, відсутність спеціалізованих навчальних наборів даних із контекстами, характерними саме для ПТСР, знижує точність таких моделей у реальних умовах застосування.

Таким чином, виникає наукове протиріччя між необхідністю комплексного та достовірного виявлення ПТСР у користувацькому контенті, особливо з урахуванням контекстної багатозначності, коморбідних станів та поведінкових проявів, і недосконалістю наявних підходів, які не враховують ці фактори у повному обсязі. Вирішення цього протиріччя вимагає розробки нових методів, що поєднують лінгвістичний, семантичний та поведінковий аналіз, з використанням адаптованих моделей штучного інтелекту та спеціалізованих наборів даних, що є науково-прикладною задачею.

Зазначена науково-прикладна задача відповідає предметній області Стандарту вищої освіти України зі спеціальності 122 Комп'ютерні науки для третього (освітньо-наукового) рівня вищої освіти, зокрема, такому об'єкту вивчення та діяльності, як «процеси збору, представлення, обробки, зберігання, передачі та доступу до інформації в комп'ютерних системах». Стандарт затверджено наказом МОН України від 28.04.2022 № 394.

Зв'язок роботи з науковими програмами, планами, темами.

Наведені в дисертації дослідження проводились в рамках виконання держбюджетних тем: 1) за Додатковою угодою № БФ/4-2025 від 01 березня 2025 року до Договору № БФ/9-2021 від 01 червня 2021 року «Розробка методів і алгоритмів ідентифікації та класифікації залишків зруйнованих будівель на основі нейромережевого аналізу візуальних даних» (ДР № 0125U003585), де автор був виконавцем; 2) «Нейромережеві методи класифікації текстильних відходів для систем циркулярної економіки» (ДР № 0126U002064), де автор був виконавцем. У межах першої теми використано

запропоновані в дисертації підходи до формування спеціалізованих навчальних наборів даних, адаптації нейромережових моделей до особливостей предметної області та оцінювання якості моделей за сукупністю метрик; у межах другої – підходи до побудови моделей класифікації, розширення множини інформативних ознак об'єктів аналізу та програмні модулі оцінювання й візуалізації результатів класифікації.

Мета і завдання дослідження. *Об'єкт дослідження:* процес виявлення ознак посттравматичного стресового розладу, супутніх психічних розладів та ризиків соціально небезпечної поведінки громадян за аналізом користувацького текстового контенту.

Предмет дослідження: методи та засоби інтелектуального аналізу користувацького контенту, обробки природної мови та нейромережового моделювання для виявлення посттравматичного стресового розладу, психічних розладів і поведінкових ризиків.

Метою дослідження є підвищення точності виявлення ознак посттравматичного стресового розладу і супутніх психічних розладів та розширення можливостей оцінювання ризиків соціально небезпечної поведінки громадян за інтелектуальним аналізом користувацького контенту засобами контекстно-орієнтованих нейромережових моделей і розширення простору ознак.

Під точністю виявлення розуміються показники класифікації на рівні повідомлення: Accuracy, Precision, Recall, F1-score. Розширення можливостей оцінювання ризиків полягає в переході від рішень щодо окремих повідомлень до скринінгового часового індексу ризику на рівні користувача, який інтегрує оцінки прояву ПТСР, інших психічних розладів та емоційного контексту повідомлень. Супутніми названо психічні розлади, ознаки яких можуть виявлятися в текстах одночасно з ознаками ПТСР; їх зв'язок саме з ПТСР на використаних наборах даних не перевірявся.

Для досягнення поставленої мети ставляться такі завдання:

1. Провести аналіз сучасних методів, моделей і засобів автоматизованого виявлення посттравматичного стресового розладу та інших психічних розладів за текстовим користувацьким контентом.

2. Удосконалити метод аналізу прояву посттравматичного стресового розладу в користувацькому контенті з урахуванням контекстних залежностей, специфічних для ПТСР.

3. Удосконалити метод виявлення психічних розладів за аналізом користувацького контенту з можливістю ідентифікації кількох розладів одночасно.

4. Розробити метод розширення множини ознак користувацького контенту для попередження соціально небезпечної поведінки громадян із ПТСР.

5. Розробити програмну систему інтелектуального аналізу користувацького контенту та експериментально дослідити ефективність запропонованих методів.

Методи дослідження. Для розв'язання поставлених завдань використано методи обробки природної мови, інтелектуального аналізу даних, машинного та глибокого навчання, трансформерні архітектури, методи багатоміткової класифікації, методи статистичного аналізу й оцінювання якості класифікаційних моделей, а також засоби візуальної інтерпретації результатів нейромережевого аналізу.

Наукова новизна отриманих результатів полягає у одержанні таких наукових результатів:

удосконалено:

1) метод аналізу прояву посттравматичного стресового розладу в користувацькому контенті, який відрізняється від існуючих формуванням спеціалізованого навчального набору даних, нецільовий клас якого містить тексти тематично суміжних спільнот, зокрема з проявами інших психічних розладів, у поєднанні з контекстно-позиційним представленням тексту, що дало змогу підвищити точність виявлення ознак ПТСР.

2) метод виявлення психічних розладів за аналізом користувацького контенту, який відрізняється від існуючих декомпозицією задачі на множини незалежних бінарних класифікаторів з окремо сформованими навчальними вибірками та індивідуальними порогами прийняття рішення, що не потребує вибору одного домінуючого розладу і не знижує точність порівняно з багатокласовою класифікацією.

вперше розроблено:

3) метод розширення множини ознак користувацького контенту для попередження соціально небезпечної поведінки громадян із посттравматичним стресовим розладом, який відрізняється від існуючих підходів інтеграцією оцінок прояву ПТСР, інших психічних розладів та емоційного контексту повідомлень у пояснюваний часовий індекс ризику на рівні користувача, що дає змогу формувати скринінгову оцінку ризику соціально небезпечної поведінки.

Практичне значення отриманих результатів. Практичне значення отриманих результатів полягає в доведенні теоретичних результатів дисертаційної роботи до реалізації та у безпосередньому використанні їх на підприємстві.

Реалізована система комплексного підходу до виявлення та класифікації проявів посттравматичного стресового розладу у користувацькому контенті за допомогою засобів штучного інтелекту. Запропоноване рішення забезпечує автоматизований аналіз цифрової активності користувачів, що охоплює модельну оцінку прояву ознак ПТСР у текстах, ідентифікацію типових патернів мовлення та поведінки з числовими характеристиками виразності, а також виявлення супутніх психічних станів. Система враховує контекстну семантику та позиційні залежності слів, що покращує точність і гнучкість аналізу. Програмна реалізація також містить модуль візуалізації результатів, що сприяє кращому розумінню роботи алгоритмів, підвищенню прозорості системи та довіри з боку фахівців і кінцевих користувачів.

Особистий внесок здобувача та внесок інших співавторів у спільних публікаціях.

Усі наукові результати дисертаційного дослідження, що виносяться на захист, отримані автором особисто. Список опублікованих праць за темою дисертації представлено в списку використаних джерел – [158–169]. Автором самостійно проведено аналіз сучасного стану досліджень у сфері автоматизованого виявлення посттравматичного стресового розладу та інших психічних розладів за користувацьким контентом, сформульовано мету, завдання та основні положення дослідження, удосконалено метод аналізу прояву посттравматичного стресового розладу в користувацькому контенті, удосконалено метод виявлення психічних розладів за аналізом користувацького контенту, розроблено метод розширення множини ознак користувацького контенту для попередження соціально небезпечної поведінки громадян із ПТСР, здійснено програмну реалізацію запропонованих методів, виконано експериментальні дослідження, проведено аналіз та узагальнення отриманих результатів.

У спільних публікаціях автору належать такі результати: метод аналізу прояву посттравматичного стресового розладу в користувацькому контенті на основі контекстно-позиційного представлення тексту та результати його експериментального дослідження [158]; принцип формування спеціалізованого помилко-орієнтованого навчального набору даних, нецільовий клас якого містить тексти з проявами інших психічних розладів, і нейромережева модель діагностування проявів ПТСР у текстовому контенті [160, 164, 165]; метод виявлення психічних розладів за аналізом користувацького контенту на основі роздільного підходу до класифікації, що забезпечує ідентифікацію кількох розладів одночасно, та багатоміткова трансформерна модель виявлення прихованих психічних станів [159, 161]; метод розширення множини ознак користувацького контенту, часовий індекс ризику та підхід до скринінгового оцінювання ризику соціально небезпечної поведінки громадян із ПТСР [162, 163, 166]; узагальнена методологія

ідентифікації індикаторів ПТСР у текстових даних [169]; архітектура, алгоритмічне забезпечення та програмна реалізація інтелектуальних інформаційних систем аналізу прояву ПТСР і виявлення психічних розладів у текстовій інформації [167, 168]. У всіх зазначених працях автором також сформовано й підготовлено навчальні набори даних, проведено обчислювальні експерименти, виконано оцінювання якості моделей та інтерпретацію одержаних результатів, підготовлено початковий варіант рукопису.

Особистий внесок інших співавторів у спільних публікаціях: у працях [160–162, 164–166, 169] О. Мазурець здійснював концептуалізацію та загальне керування дослідженням, а також рецензування й коригування чорнетки рукопису; у свідоцтвах про реєстрацію авторського права [167, 168] О. Мазурцю належать загальна концепція програмних систем і керування їх розробленням; у статті [158] І. Крак та О. Бармак брали участь у постановці наукової проблеми та загальному обговоренні отриманих результатів, О. Мазурець здійснював концептуалізацію, керування дослідженням і коригування чорнетки рукопису, М. Молчанова, О. Собко та О. Тищенко спільно з автором виконували підготовку даних, перевірку та візуалізацію результатів експериментів; у статті [159] І. Крак та О. Бармак брали участь у постановці наукової проблеми та загальному обговоренні отриманих результатів, О. Мазурець здійснював концептуалізацію, керування дослідженням і коригування чорнетки рукопису, М. Молчанова спільно з автором виконувала підготовку даних і візуалізацію результатів, Л. Азарова здійснювала лінгвістичний аналіз мовних маркерів і редагування рукопису; у статті [163] І. Крак та О. Бармак брали участь у постановці наукової проблеми та загальному обговоренні отриманих результатів, О. Мазурець здійснював концептуалізацію, керування дослідженням і коригування чорнетки рукопису, О. Собко та М. Молчанова спільно з автором виконували підготовку даних, перевірку та візуалізацію результатів експериментів. Ідеї,

положення та результати, що належать співавторам, у дисертації використано з відповідними посиланнями.

Апробація матеріалів дисертації. Основні результати дисертаційного дослідження доповідались та обговорювались на міжнародних і всеукраїнських науково-технічних та науково-практичних конференціях, зокрема: CLW-2025: Computational Linguistics Workshop at the 9th International Conference on Computational Linguistics and Intelligent Systems «CoLInS-2025» (May 15-16, 2025, Kharkiv, 2025), Modern Machine Learning Technologies Workshop «MoMLLeT-2025» (June 14-15, 2025, Lviv, Ukraine), 5th International Workshop of IT-professionals on Artificial Intelligence «ProfIT AI 2025» (October 15-17, 2025, Liverpool, United Kingdom), XIII Всеукраїнська науково-практична конференція «Глушковські читання» (6 грудня 2024, Київ), XXIII Міжнародна наукова конференція «Нейромережні технології та їх застосування НМТіЗ-2024» (11-12 грудня 2024, Краматорськ-Тернопіль), XXIII Міжнародна науково-практична конференція «Математичне та програмне забезпечення інтелектуальних систем МПЗІС-2025» (19-21 листопада 2025, Дніпро).

Публікації. Основні результати дисертації висвітлено у 12 наукових працях (Додаток А), з яких 3 статті у наукових виданнях, включених до Переліку наукових фахових видань України, 6 публікацій, які засвідчують апробацію матеріалів дисертації у матеріалах конференцій, а також 3 публікації, які додатково відображають наукові результати дисертації.

Структура та обсяг дисертації. Дисертація складається з анотації, змісту, переліку умовних скорочень, вступу, чотирьох розділів, висновків, списку використаних джерел із 169 найменувань на 24 сторінках та 4-х додатків на 10 сторінках. Загальний обсяг дисертаційної роботи становить 188 сторінок друкованого тексту, з них 135 сторінок основного тексту. Дисертація містить 33 рисунки та 13 таблиць.

РОЗДІЛ 1. АНАЛІЗ МЕТОДІВ, ЗАСОБІВ ТА ТЕХНОЛОГІЙ ДЛЯ АВТОМАТИЗОВАНОГО ВИЯВЛЕННЯ ПТСР У КОРИСТУВАЦЬКОМУ ТЕКСТОВОМУ КОНТЕНТІ

1.1. Сучасний стан інтелектуального аналізу текстового контенту для автоматизованого виявлення ПТСР

Посттравматичний стресовий розлад (ПТСР) є виснажливим психічним розладом, на який страждають приблизно 10 % населення [24, 25]. З розвитком цифрових технологій і соціальних медіа з'явилася нова можливість для вивчення проявів ПТСР через аналіз користувацького контенту [26]. Ця тема набуває особливої актуальності в контексті збройних конфліктів, природних катастроф і інших травматичних подій, які часто залишають глибокий відбиток на психіці людей.

Аналіз користувацького контенту, такого як повідомлення в соціальних мережах, блоги, форуми і коментарі, дозволяє отримати унікальні інсайти щодо емоційного і психічного стану постраждалих [27]. Застосування методів машинного навчання та обробки природної мови для виявлення ознак ПТСР у текстах користувачів є сучасним і перспективним напрямом наукових досліджень, адже можуть не тільки ідентифікувати окремі випадки ПТСР, але й допомогти в розробці превентивних заходів і програм підтримки для осіб, які перебувають у зоні ризику.

В умовах сучасного суспільства, яке зазнає впливу військових дій, пандемій тощо аналіз користувацького контенту для виявлення ПТСР є надзвичайно важливим. Це дає можливість не лише науковцям, а й практикам в галузі психології та соціальної роботи оперативно реагувати на потреби населення [27].

Протягом багатьох років наукове співтовариство досліджувало підходи до моніторингу для виявлення певних психічних розладів і ризикової поведінки, таких як депресія, розлади харчової поведінки, азартні ігри та

суїцидальні думки, серед іншого, щоб активувати стратегії профілактики або пом'якшення, а у важких випадках – клінічні лікування [28]. Показники посттравматичного стресового розладу значно зросли через пандемію COVID-19 [29]. Частково це пов'язано із ізолюваністю або недоступністю терапевтичного втручання, викликаного пандемією. Додаткові інструменти скринінгу можуть знадобитися для посилення ідентифікації та діагностики посттравматичного стресового розладу за допомогою віртуального середовища. У дослідженні [17] пропонується ідентифікувати людей з посттравматичним стресовим розладом за допомогою аналізу настроїв із напівструктурованих інтерв'ю. Для цього використано модель машинного навчання, яка навчається на текстових даних, які є частиною корпусу аудіо/візуальних викликів емоцій і семінару (AVEC-19 [30]). Розмір вибірки становить 188 осіб без посттравматичного стресового розладу та 87 із посттравматичним стресовим розладом. Модель Random Forests змогла досягти збалансованої точності в 80,4 % на масиві даних, використаному під час тестування AVEC-19. Схожий підхід був у [31] використаний підхід машинного навчання, щоб визначити відносну важливість конкретних труднощів регуляції емоцій щодо тяжкості симптомів ПТСР.

Великі досягнення в області виявлення ПТСР також належить великим мовним моделям. ChatGPT продемонстрував попередню початкову здійсненність для цієї мети, однак, чи може він точно оцінити психічні захворювання, було досліджено у роботах [19, 32]. Порівнювалась ефективність ChatGPT і моделі text-embedding-ada-002 (ADA) у виявленні посттравматичного стресового розладу після пологів (CB-PTSD), післяпологового психічного захворювання матері, яке вражає мільйони жінок щорічно, без стандартного скринінгу. Використовуючи вибірку з 1295 жінок, які народили за останні шість місяців і яким було 18 років і старше, набраних через оголошення лікарень, соціальні мережі та професійні організації, було досліджено потенціал ChatGPT і ADA для скринінгу CB-PTSD шляхом аналізу розповідей про пологи у матері. Контрольний список

посттравматичних стресових розладів для DSM-5 [33] використовувався для оцінки CB-PTSD. Запропонована модель ML, яка використовує числове векторне представлення моделі ADA, було ідентифіковано CB-PTSD за допомогою наративної класифікації. Запропонована модель перевершила (оцінка F1: 0,81) ChatGPT і шість раніше опублікованих великих моделей вбудовування тексту, навчених на даних психічного здоров'я або клінічних сфер, що свідчить про те, що модель ADA можна використовувати для виявлення CB-PTSD. Запропонований підхід до моделювання можна узагальнити для оцінки інших розладів психічного здоров'я.

У дослідженні [18] досліджується мова як потенційний діагностичний біомаркер ПТСР. Проаналізовано оригінальну вибірку зі 148 осіб, які стали жертвами терактів у Парижі 13 листопада 2015 року. Інтерв'ю, проведені через 5–11 місяців після події, охоплюють осіб зі схожим соціально-економічним становищем, які зазнали того самого інциденту, відповідаючи на ідентичні запитання та використовуючи однакові заходи щодо ПТСР. Використовуючи цей набір даних для збору нюансів, які можуть бути клінічно значущими, запропоновано триетапну міждисциплінарну методологію, яка об'єднує досвід психіатрії, лінгвістики та спільноти обробки природної мови для вивчення зв'язку між мовою та ПТСР. На першому етапі оцінюється здатність клінічного психіатра діагностувати посттравматичний стресовий розлад, використовуючи лише транскрибування інтерв'ю. На другому етапі використовується статистичний аналіз і моделі машинного навчання для створення мовних особливостей на основі психолінгвістичних гіпотез і оцінки їх передбачуваної сили. Третій крок застосування підходу глибокого навчання без гіпотез до класифікації ПТСР у обраній вибірці. Результати показують, що клінічний психіатр поставив діагноз ПТСР з AUC 0,72. Це можна порівняти з анкетною золотого стандарту (площа під кривою (AUC) $\approx 0,80$). Модель машинного навчання досягла діагностичного AUC 0,69. Підхід глибокого навчання досяг AUC 0,64. Важливо, що в дослідженні контролюються фактори, що змішують, встановлюються зв'язки між мовою та субсимптомами DSM-5, а також

поєднуються автоматизовані методи з якісним аналізом. Це дослідження забезпечує прямий і методологічно надійний опис зв'язку між ПТСР і мовою. Схоже дослідження проводилось для військового персоналу, розгорнутого в Афганістані. AUC моделі, розробленої в цьому дослідженні, становив 0,74 [34].

Було протестовано п'ять керованих алгоритмів машинного навчання (Elastic Net, Gradient Boosting Machine, Random Forest, Support Vector Machine і C5.0) з використанням клінічних даних (версія 5 шкали посттравматичних стресових розладів, керованих лікарем) та соціально-демографічних характеристик [35]. У дослідженні взяли участь 112 пацієнтів (61 з NET-Trauma та 51 з PROVE). Найкращі результати для діагностики посттравматичного стресового розладу показала модель з чотирма класами на основі алгоритму C5.0, яка використовувала дані CAPS-5 з 15 елементами та соціально-демографічні ознаки. Точність цієї моделі склала 65,6 % для навчального набору даних та 52,9 % для тестового набору. Основними характеристиками для прогнозування стадії ПТСР були кількість симптомів, загальна оцінка CAPS-5, загальна оцінка тяжкості та наявність поточних або минулих травматичних подій.

За результатом аналізу пов'язаних робіт у сфері аналізу прояву посттравматичного стресового розладу в користувацькому контенті, виявлено, що відомі рішення побудовано переважно на клінічних інтерв'ю та наративах вузьких груп і вони не передбачають контрольованого відмежування ознак ПТСР від тематично суміжних станів на рівні дописів соціальних мереж. Наведені в них показники (збалансована точність 80,4 % [17], F1 0,81 [19], AUC 0,64–0,74 [18, 34]) отримано в різних постановках і на різних даних, тому вони не є безпосередньо зіставними. Тому модель ІІІ, яка використовується для задачі виявлення ПТСР, повинна:

- мати високу здатність до генералізації, що повинно дозволити їй краще впоратися з новими або невідомими даними, оскільки різні люди можуть висловлювати свої переживання по-різному;

- бути контекстно-орієнтованою (для виявлення специфічних мовних ознак, які можуть свідчити про наявність ПТСР);
- надавати можливість візуальної інтерпретації одержаних моделлю результатів.

Отже, попри наявність значної кількості наукових праць і реалізованих моделей, актуальною залишається проблема досягнення високої точності та інтерпретованості результатів. Існуючі підходи демонструють обмеженість у контекстному розумінні, генералізації та ідентифікації латентних патернів ПТСР у текстах, що ускладнює їхнє застосування в умовах реального середовища. Отже, виникає потреба в удосконаленні методів штучного інтелекту, здатних не лише виявляти симптоми ПТСР з урахуванням контексту, а й надавати прозорі пояснення прийнятих рішень для подальшого використання в психологічній та соціальній практиці.

1.2. Сучасний стан виявлення психічних розладів за текстовим контентом

Психічні розлади суттєво впливають на якість життя та часто залишаються не діагностованими. У [36] досліджується можливість раннього виявлення психічних розладів через моніторинг мови в соціальних мережах за допомогою моделей глибокого навчання, зокрема для розпізнавання депресії, анорексії та схильності до самоушкодження. Моделі доповнюються обчислювальним аналізом, заснованим на теоріях психології для кращого розуміння когнітивних стилів та емоцій.

У [37] проаналізовано лінгвістичні особливості постів користувачів на субреддитах, пов'язаних із психічними розладами, та порівняно їх із постами на субреддитах, не пов'язаних із психічним здоров'ям. Використовуючи методи статистичного аналізу та машинного навчання, зокрема BERT, було виявлено значні відмінності в лексичних та семантичних паттернах. Це підтверджує можливість використання мовних

ознак для раннього виявлення психічних розладів, що може допомогти клініцистам у діагностиці та онлайн-втручанні.

Ряд дослідників підтвердили ефективність виявлення суїцидальних думок через аналіз онлайн-активності та соціальних даних [38]. Наприклад, у [39] досліджено важливість емоцій у вмісті постів Twitter для виявлення ризиків суїциду за допомогою SentiStrength і NRC Affect Intensity Lexicon. Автори [40] використали LIWC для автоматичної ідентифікації передсмертних записок із публікацій у соціальних мережах, навчаючи модель LSTM. У [41] досліджено аналіз n-грам і виявлено, що фрази, пов'язані з суїцидальними тенденціями, часто зустрічаються на форумах про самогубство. Найефективнішою виявилася комбінована модель LSTM-CNN, яка досягла 93,8 % точності та 92,8 % результату F1 для бінарної класифікації.

У [42] досліджувались лінгвістичні показники психічного здоров'я, які охоплювали і питання психічного здоров'я. Автори використовували BERT для класифікації серед восьми класів, але вони досягли середньої точності приблизно 64 % і середнього балу F1 близько 0,645, що не є високим результатом для проблеми психічного здоров'я. Ще одне схоже дослідження [43] у напрямку багатокласової класифікації розглядало використання методів штучного інтелекту для виявлення психічних розладів, таких як депресія [44], тривожність [45, 46], біполярний розлад [47, 48], ADHD [49, 50] та ПТСР, за допомогою аналізу даних користувачів Reddit [51]. Пропонована модель на основі трансформерів з методом пізнього злиття (late fusion) для обробки заголовків та постів показала кращі результати (F1 0,8965) порівняно з іншими традиційними моделями.

Автори дослідження [52] зробили огляд застосування методів машинного та глибокого навчання для прогнозування стану психічного здоров'я серед студентів з 2011 по 2024 рік. Виявлено, що згорткові нейронні мережі, випадковий ліс, опорна векторна машина та інші моделі показали значні результати, зокрема CNN, SVM, ELM та нейронні мережі, ефективно

прогнозують психічні розлади на основі демографічних даних, клінічних оцінок, контенту соціальних медіа та нейровізуалізації. CNN досягли найвищої точності в діагностиці біполярного розладу (99,75 %), тоді як SVM показали 95 % для AUC у випадку шизофренії, а ELM - найвищий AUC для ADHD (0,8757). Однак дослідження підкреслює необхідність у більш широких і різноманітних наборах даних та врахуванні тимчасової динаміки для подальшого покращення діагностики.

Дослідження [53] демонструє ефективність аналізу даних із соціальних мереж для виявлення симптомів депресії, зокрема через Twitter. Дослідники аналізували чотири різні BERT-моделі: DBUFS2E, BBU, MBBU та DRB. Вони порівняли їх із базовою моделлю (Baseline) та традиційними методами машинного навчання, такими як CatBoost і GBC. Результати дослідження показують, що моделі DBUFS2E і BBU досягли найвищої точності 97 %, F1-метрики 96 %–97 % та AUC 98 %, перевершуючи традиційні методи. MBBU і DRB також показали високі результати Accuracy 95 %–96 %. Baseline 75 % значно поступається, а Catboost і GBC Accuracy 91 % менш ефективні за BERT-моделі, що підтверджує перевагу трансформерів у виявленні симптомів депресії. Також результати вказують, що доповнення текстових даних іншими джерелами, такими як емодзі, зображення та хештеги, може покращити прогнозування.

Автори дослідження [54] проводять класифікацію психічних розладів на основі текстових даних із Reddit, застосовуючи Multi-Class Multi-Level (MCML) підхід, який перевершує традиційний Multi-Class Single-Level (MCSL) алгоритм. Використано методи машинного, глибокого та трансферного навчання, а також два сценарії попередньої обробки тексту. Найкращі результати досягнуто за допомогою RoBERTa, яка показала точність 98 % на першому рівні класифікації та 85 % на другому. MCML підтверджує свою ефективність для раннього виявлення психічних розладів на основі текстового аналізу.

У дослідженні [55] автори демонструють, що традиційні методи (BOW, TF-IDF, Glove) та загальні мовні моделі (BERT [56, 61], RoBERTa [57], DistilBERT [58, 132]) мають обмежену ефективність у багатокласовій класифікації психічних розладів через слабе розуміння контексту та метафоричних виразів. Для вирішення цієї проблеми запропоновано гібридну трансформерну архітектуру, що поєднує MentalBERT [59] (навчений на текстах про психічне здоров'я) та MelBERT [60] (спеціалізований на метафорах) з CNN для глибшої обробки тексту. Запропонована модель досягла точності 92 %, перевершуючи результати моделей, що були наведені у дослідженні.

Аналіз сучасних досліджень у сфері виявлення психічних розладів на основі користувацького контенту в соціальних мережах засвідчує ефективність застосування методів машинного та глибокого навчання, зокрема трансформерних архітектур. Науковці активно досліджують лінгвістичні особливості мови, емоційне забарвлення повідомлень, структурні патерни та семантичні характеристики текстів, що дозволяє виявляти симптоми психічних розладів із високим ступенем точності. Результати численних експериментів підтверджують, що моделі, адаптовані до специфіки психоемоційного контенту, демонструють значну перевагу над класичними алгоритмами. У поєднанні з психологічними теоріями та мультимодальними джерелами даних такі моделі дають змогу не лише ідентифікувати наявність розладів, а й прогнозувати ризики небажаної поведінки. Це відкриває перспективи для створення більш точних та адаптивних систем підтримки психічного здоров'я, зокрема в контексті раннього втручання та превентивних заходів.

Однак, за результатами наведеного аналізу виявлено досить низьку точність при багатокласовій класифікації, а також проблеми при ідентифікації декількох психічних розладів одночасно. Порівняльну характеристику розглянутих підходів наведено в таблиці 1.1.

Таблиця 1.1

Порівняння підходів до автоматизованого виявлення ПТСР і психічних
розладів за текстом

Дані	Модель / метод	Результат	Обмеження щодо задач роботи
Напівструктуровані інтерв'ю (корпус AVEC-19); 87 осіб із ПТСР, 188 без	Аналіз настроїв, Random Forest [17]	Збалансована точність 80,4 %	Інтерв'ю, а не користувацький контент; класичне МН без урахування контексту
Наративи про пологи, 1295 жінок (CB-PTSD)	ChatGPT; ембединги text-embedding-ada-002 + МН [19, 32]	F1 0,81 (ADA), ChatGPT гірше	Вузька група та тип травми; закрита модель, відсутність пояснень
Транскрипти інтерв'ю 148 постраждалих від терактів	Психолінгвістичні ознаки + МН; глибоке навчання без гіпотез [18]	AUC 0,69 (МН), 0,64 (DL); психіатр – 0,72	Малий обсяг; клінічні інтерв'ю; контекстні залежності не моделюються
Інтерв'ю військового персоналу	Мовні маркери + МН [34]	AUC 0,74	Та сама методологія, що й [18]
Клінічні дані CAPS-5 та соціально-демографічні ознаки, 112 пацієнтів	Elastic Net, GBM, RF, SVM, C5.0 [35]	Точність 65,6 % (навчання), 52,9 % (тест)	Не текст; низька узагальнювальна здатність
Дописи в соціальних мережах	Моделі глибокого навчання + психологічні теорії [36]	Якісна оцінка для депресії, анорексії, самоушкодження	ПТСР не розглядається
Дописи субредитів про психічне здоров'я	Статистичний аналіз + BERT [37]	Значущі лексико-семантичні відмінності	Виявлення відмінностей, а не класифікація конкретного розладу
Форуми про самогубство	LSTM-CNN, n-грами [41]	Точність 93,8 %, F1 92,8 % (бінарна)	Лише суїцидальні тенденції; бінарна постановка
Дописи Reddit, 8 класів психічних станів	BERT, багатокласова класифікація [42]	Точність $\approx$ 64 %, F1 0,645	Низька точність багатокласової постановки
Reddit: депресія, тривожність, біполярний розлад, ADHD, ПТСР	Трансформери з пізнім злиттям заголовка та посту [43]	F1 0,8965	Один домінуючий клас на текст; коморбідність не враховується
Twitter, симптоми депресії	BERT-моделі (DBUFS2E, BBU, MBBU, DRB) проти CatBoost, GBC [53]	Точність 97 %, F1 96–97 %, AUC 98 %	Лише депресія; бінарна постановка
Reddit, кілька розладів	RoBERTa, ієрархічна MCML-класифікація [54]	Точність 98 % (1-й рівень), 85 % (2-й)	Ієрархія взаємовиключних класів, а не сумісні мітки
Тексти про психічне здоров'я, багатокласова задача	MentalBERT + MeIBERT + CNN [55]	Точність 92 %	Взаємовиключні класи; складна гібридна архітектура

Класифікацію психічних розладів наведено на рис. 1.1.

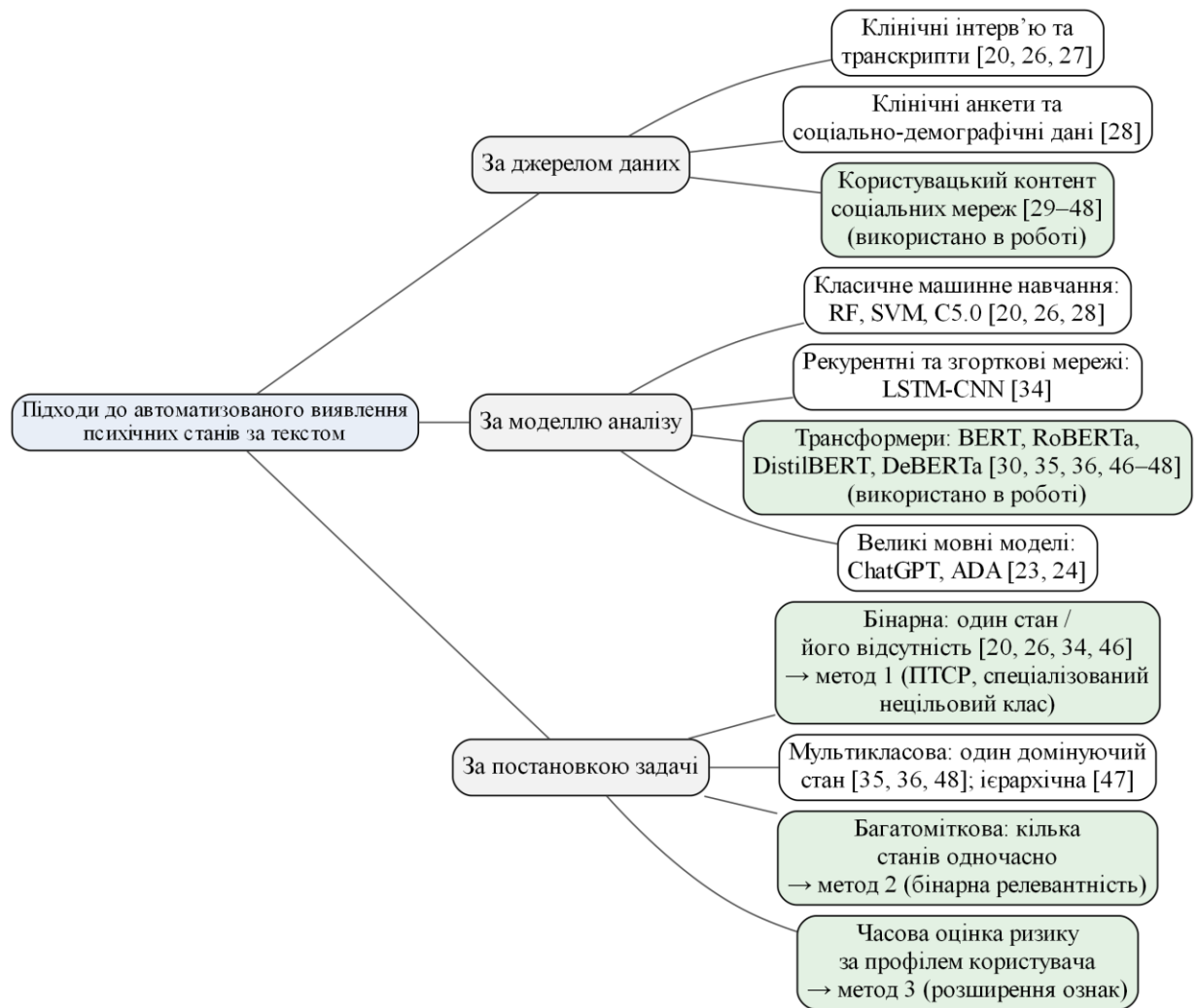


Рис. 1.1 – Класифікація підходів до виявлення психічних станів за текстовим користувацьким контентом

Проведений аналіз дає змогу виокремити три невирішені проблеми, які визначають зміст роботи: 1) існуючі моделі виявляють загальні мовні маркери психічного стану, але не контекстні залежності, специфічні для ПТСР, і не відмежовують його від суміжних станів; 2) багатокласова постановка примусово обирає один домінуючий розлад, тоді як коморбідність є типовою; 3) поведінкові ризики оцінюються за окремими лінгвістичними ознаками без урахування їх часової динаміки та результатів інших методів. Розв'язанню цих проблем присвячено відповідно методи 1, 2 і 3 (розділ 2).

1.3. Урядові програми та нормативні документи

Порушення психічного здоров'я людини негативно впливає на її фізичний стан, часто супроводжується залежностями та ускладнює соціальну інтеграцію, що в підсумку знижує працездатність і якість життя [62]. Це становить серйозну проблему не лише для окремої особи, а й для всієї держави, вимагаючи системної відповіді. Збереження психічного здоров'я стало критично важливою складовою державної політики, адже від нього залежить як ефективність функціонування Збройних Сил України, так і загальна стійкість суспільства в умовах зовнішніх загроз.

На сайті Кабінету Міністрів України висвітлено проведення установчої конференції, присвяченої формуванню спільноти фахівців для більш ефективної допомоги людям із ПТСР [63]. Обговорювалися підходи до навчання, обміну досвідом та міждисциплінарної взаємодії, що сприятиме розвитку цілісної системи підтримки. Конференція стала важливим кроком у реалізації Національної програми психічного здоров'я.

Цільова модель 2.0 системи психічного здоров'я та психосоціальної підтримки в Україні – офіційний документ, що описує стратегічні цілі, підходи, цільові групи, нормативну базу, ресурси та напрямки розвитку національної системи ПЗПСР [64].

Цільова модель 2.0 є концептуальним фундаментом для трансформації системи психічного здоров'я в Україні [65]. Вона відображає міждисциплінарний підхід, в якому медичні, соціальні, освітні та правові інституції діють скоординовано з метою забезпечення доступу до якісних, своєчасних і людиноцентричних послуг. Модель враховує особливості повоєнного стану суспільства, поширеність ПТСР та інших розладів, а також потребу в профілактиці та ранньому втручанні.

Таким чином, психічне здоров'я постає не лише як індивідуальний ресурс, а як ключовий чинник суспільної стійкості, національної безпеки та ефективного функціонування державних інституцій. Реалізація Цільової

моделі 2.0 свідчить про стратегічне переосмислення підходів до психосоціальної підтримки в Україні, засноване на міжсекторальній співпраці, децентралізації допомоги та орієнтації на потреби людини. Така трансформація має потенціал не лише подолати наслідки травматичних переживань, спричинених війною, а й сформувати стійку систему охорони психічного здоров'я, здатну відповісти на сучасні виклики та інтегруватися у європейський простір.

З нормативного контексту впливають вимоги до методів, що розробляються в дисертації: 1) орієнтація на скринінг і раннє виявлення на рівні первинної ланки допомоги, а не на встановлення діагнозу; 2) робота зі знеособленими даними та дотримання вимог захисту персональних даних; 3) пояснюваність результатів для фахівця, який ухвалює рішення; 4) зниження ризику стигматизації, тобто контроль хибнопозитивних рішень. Ці вимоги враховано в постановці задачі (розділ 2) та в п. 3.8.

1.4. Аналіз ознак психічних розладів у текстовому контенті

Клінічний протокол МОЗ України [66] трактує ПТСР як тривалий і глибокий психічний стан, що виникає внаслідок переживання людиною надзвичайно травматичної ситуації, яка перевищує її здатність до адаптації [67]. У контексті психології ПТСР [68] вважається відповіддю на загрозу життю, серйозні фізичні ушкодження чи психологічне потрясіння, що порушує базове відчуття безпеки та стабільності. Діагностичні критерії ПТСР визначено у DSM-5-TR [33].

Не кожна ситуація, яка викликає емоційне напруження чи тривогу, призводить до розвитку цього розладу. Стреси, зумовлені повсякденними викликами – такими як професійне виснаження, міжособистісні конфлікти чи побутові труднощі, – хоч і можуть впливати на психічне здоров'я, не мають тієї руйнівної сили, яка є характерною для подій, що запускають механізми формування ПТСР [69]. Йдеться про якісно інший рівень психологічного

навантаження, коли психіка втрачає можливість самостійно відновити внутрішню рівновагу без зовнішнього втручання.

Науковий підхід до діагностики ПТСР враховує не лише сам факт наявності стресу, а й інтенсивність, тривалість і специфіку реакцій особистості на нього. Ключову роль відіграють феномени повторного переживання травми, уникання, гіперзбудження та когнітивних порушень, які можуть зберігатися місяцями або навіть роками після інциденту. Таким чином, ПТСР – це не просто реакція на складні обставини, а складне психологічне порушення, яке потребує професійної допомоги [70].

Клінічна картина ПТСР [71] охоплює низку психопатологічних проявів, серед яких характерними є уникання травматичних нагадувань, підвищене збудження нервової системи, повторне переживання травматичних епізодів (так званий флешбек), а також порушення когнітивного функціонування й емоційної регуляції.

Клінічні настанови та огляди [72, 73] наголошують, що розвиток ПТСР не завжди є миттєвою реакцією на стресову подію. Іноді симптоми проявляються після відносно тривалого латентного періоду – від кількох тижнів до кількох місяців. На етапі стабілізації, коли зовнішні обставини перестають бути загрозливими, можуть активізуватися глибинні психічні механізми, що сприяють формуванню патологічної стресової відповіді. Це свідчить про складну, багатоетапну природу ПТСР, який може залишатися непомітним у періоді негайного постстресу, але проявлятися згодом у формі дезадаптивних психологічних реакцій.

Наприклад, у [74] здійснено аналіз еволюції діагностичних підходів до ПТСР, із урахуванням змін у клінічних критеріях та контекстуальних особливостях його прояву. Автори досліджують епідеміологічні закономірності поширення цього розладу, зокрема акцентуючи увагу на соціокультурній варіативності досвіду травми. Особлива увага приділяється виникненню посттравматичних симптомів у ситуаціях, які традиційно не підпадали під визначення травматичних, що, у свою чергу, ставить під сумнів сталі межі діагностичної категорії.

Розуміння динаміки ПТСР та чинників, які впливають на процес відновлення, має критичне значення для своєчасного виявлення випадків, коли психосоціальні втручання можуть бути найбільш ефективними. У [75] проаналізовано дані численних обсерваційних досліджень, з метою виявлення закономірностей у перебігу ПТСР протягом першого року після травматичної події. Методологія ґрунтувалася на метааналітичному підході, що дозволив простежити зміну рівня поширеності ПТСР на основі щонайменше дворазового оцінювання одного й того ж контингенту впродовж року після травми, а також – у деяких випадках – до двох років. У дослідженні було вивчено, яким чином перебіг ПТСР може модифікуватися в залежності від типу травматичної події, способу діагностичного оцінювання, демографічних характеристик вибірки, зокрема віку та статі. Особливу увагу приділено розрізненню між навмисними та ненавмисними травматичними подіями, що дозволяє глибше зрозуміти механізми стійкості або вразливості до розвитку розладу. Оцінка здійснювалася як за допомогою клінічного інтерв'ювання, так і на основі самозвітів, що надало змогу врахувати можливі відмінності у результатах, пов'язані з методом збору даних.

Аналіз ознак психічних розладів у текстовому контенті є важливою складовою сучасної психології, оскільки дозволяє виявляти потенційні ознаки розладів на основі текстових даних, зокрема повідомлень, постів чи коментарів. Вивчення текстів з психологічної точки зору дає змогу виявляти специфічні патерни мовлення, що можуть бути пов'язані з різними психічними станами, такими як депресія, тривожність, ПТСР або суїцидальні нахили.

Технічний аспект цього процесу включає застосування методів обробки природної мови [76], зокрема лексичний та семантичний аналіз [77], що дозволяють не тільки виявляти характерні лексичні конструкції, а й досліджувати контекст, емоційний окрас [78, 79] і настрій тексту. Використання таких інструментів, як аналіз тональності, виявлення сутностей [80], класифікація за емоційним тоном та застосування методів машинного навчання для виявлення шаблонів у мовленні, дає змогу створити

ефективні системи для моніторингу психологічного стану осіб через їх текстовий контент.

Один з ключових аспектів цього аналізу – це точність і чутливість алгоритмів, здатних виявляти психічні розлади в ранніх стадіях, що може бути особливо важливим у випадках, коли особа не звертається до фахівців, але надає певні сигнали через текст. Прогрес у цій галузі вимагає інтеграції різноманітних методів, включаючи глибоке навчання, для забезпечення точного виявлення психологічних аномалій. Відповідність клінічних проявів ПТСР, їх мовних маркерів та інформаційних ознак, придатних для автоматизованого аналізу, наведено в таблиці 1.2.

Таблиця 1.2

Відповідність клінічних проявів ПТСР, мовних маркерів та інформаційних ознак

Клінічний прояв ПТСР	Мовний маркер у користувацькому контенті	Інформаційна ознака	Де використано
Повторне переживання травми (флешбеки, нав'язливі спогади)	Лексика травматичних подій, повторюваність тем, опис у минулому часі	Контекстні залежності між токенами, семантичне представлення речень	Метод 1
Уникання, соціальна ізоляція	Заперечувальні конструкції, лексика самотності, зниження активності	Динаміка кількості дописів і тем у часі	Метод 3
Гіперзбудження, дратівливість, спалахи гніву	Агресивна та погрозна лексика, окличні конструкції	Рівень агресії, наявність небезпечної лексики (табл. 2.1)	Методи 2, 3
Негативні зміни настрою й когніцій (безнадійність, провина, страх)	Негативно забарвлена лексика, вираження страху, смутку, відрази	Тональність тексту, емоції за NRC, індикатор негативності	Методи 2, 3
Коморбідні стани (тривога, депресія, панічні прояви)	Специфічні маркери кожного розладу в одному тексті	Вектор оцінок бінарних класифікаторів розладів	Метод 2
Когнітивна дезорганізація, емоційна нестабільність	Граматичні помилки, різкі перепади тональності між реченнями, зміна використання емодзі	Кількість помилок, перепади тональності, кількість смайлів (табл. 2.1)	Метод 3

Таким чином, ПТСР постає як складне психічне явище, що формується внаслідок глибокого емоційного потрясіння, перевищуючи адаптаційні можливості особистості. Його діагностика виходить за межі виявлення реакції на стрес і потребує врахування специфіки когнітивних, емоційних та фізіологічних проявів, які можуть мати пролонгований характер. Розуміння багатофакторної природи ПТСР, включно з соціокультурним контекстом, латентною динамікою симптомів та особливостями реакцій на різні типи травматичних подій, є необхідною умовою для ефективного втручання та підтримки осіб, які пережили психотравмувальні обставини. У цьому контексті, зростаюче значення набуває аналіз текстового контенту як інструмент раннього виявлення психологічних порушень. Поєднання методів психолінгвістики та алгоритмів обробки природної мови відкриває нові можливості для моніторингу психічного здоров'я, підвищення чутливості діагностичних систем та розвитку превентивних стратегій у сфері ментального благополуччя.

1.5. Аналіз ознак користувацького контенту для попередження соціально небезпечної поведінки

У рамках дослідження, спрямованого на раннє виявлення ризиків соціально небезпечної поведінки [81, 82] серед осіб із симптомами ПТСР, особливу увагу приділено аналізу користувацького контенту, зокрема текстових повідомлень у соціальних мережах [83], чатах, блогах тощо. Тексти, які генеруються такими користувачами, можуть містити непрямі індикатори емоційного стану, психологічної нестабільності [84, 85], агресії [86] чи апатії, що робить їх джерелом цінної інформації для оцінки потенційних ризиків [87].

Для забезпечення ефективного аналізу таких текстів запропоновано розширену множину ознак, які дозволяють здійснити багатовимірну оцінку стану користувача. Основні з них охоплюють як емоційні, так і поведінкові

характеристики, що виявляються через мовні конструкції, лексичні патерни та структурні особливості тексту.

Ознаки користувацького контенту, що використовуються для попередження соціально небезпечної поведінки громадян із ПТСР, становлять собою багаторівневу систему характеристик [88-90], яка дозволяє здійснити глибокий психоемоційний аналіз текстових повідомлень [91]. Такі ознаки не зводяться до простого виявлення негативної тональності – вони охоплюють широкий спектр мовних, стилістичних та емоційних параметрів, що в комплексі здатні вказати на потенційно ризикову поведінку.

Одним із базових рівнів аналізу є виявлення загального емоційного забарвлення тексту. Загальна тональність слугує первинним фільтром, що дозволяє визначити, чи переважає у висловлюваннях позитивне, негативне чи нейтральне ставлення. Водночас набагато більшу діагностичну цінність має аналіз локальних змін емоційного стану, який здійснюється через визначення тональності окремих речень. Раптові переходи від позитивних до негативних висловлювань (або навпаки) можуть свідчити про нестабільність емоційного стану, прояви внутрішнього конфлікту чи емоційне вигорання.

У неформальному користувацькому контенті важливими маркерами є також невербальні елементи, зокрема смайли [92]. Їхня присутність або відсутність, частота використання та контекст можуть бути свідченням намагання приховати або навпаки – продемонструвати свій емоційний стан. Крім того, зростання кількості граматичних помилок у повідомленнях розглядається як індикатор психологічної напруги [93] або когнітивного перевантаження: у таких станах користувачі часто нехтують формальними правилами мови.

Окремо слід виділити рівень агресії [94] в тексті. Його виявляють за допомогою словникових ресурсів або глибоких нейронних моделей, здатних розпізнавати як явну, так і латентну агресію. Ця ознака є критично важливою при виявленні потенційно небезпечної поведінки, оскільки агресивні

висловлювання можуть бути спрямовані як назовні (до інших осіб чи груп), так і всередину – у вигляді автоагресії [95].

Для побудови детальної емоційної картини аналізуються також конкретні емоції, виражені в тексті. Наприклад, використання лексикону NRC Emotion Lexicon [96] дозволяє виокремити такі базові емоції, як страх, гнів, смуток, відраза, довіра тощо. Не менш важливим є аналіз їхньої динаміки – посилення негативних емоційних проявів, поєднане з відсутністю позитивних емоцій, може вказувати на стан глибокої апатії або депресивні розлади.

Зміни в емоційному забарвленні тексту на рівні окремих речень – зокрема, їхня різка поляризація – іноді є маркером втрати самоконтролю, емоційної дестабілізації або нав'язливих думок. Подібні патерни можуть бути проявами кризового стану, особливо якщо паралельно виявляються загрози, прямі або завуальовані, а також лексика, що натякає на можливість завдання шкоди собі чи іншим.

Не менш значущим показником є переважання негативних суджень і висловлювань у тексті. Їх накопичення може відображати тривалу фрустрацію [97], наявність емоційних тригерів, незадоволеність базовими потребами, відчуття безпорадності або соціальної ізоляції – усе це суттєво підвищує ризики розвитку соціально небезпечної поведінки.

Таким чином, комплексний аналіз користувацького тексту – з урахуванням не лише тональності, а й емоційної структури, стилістичних особливостей, агресивності та стабільності емоцій – дає змогу створити ефективну модель для виявлення громадян, які перебувають у стані ризику, ще до того, як їхня поведінка стане загрозливою для себе або оточення.

1.6. Пояснюваність результатів

У сучасному етапі розвитку штучного інтелекту спостерігається стрімке впровадження алгоритмів машинного навчання у критично важливі сфери діяльності: охорона здоров'я, фінанси, оборона, правосуддя, кібербезпека тощо.

Попри безпрецедентну ефективність таких систем, одним із найскладніших викликів залишається питання прозорості й пояснюваності їхніх рішень. Це особливо актуально в контексті моделей, побудованих на базі глибокого навчання (deep learning) [98, 99], які, незважаючи на високу точність, зазвичай функціонують як «чорні скриньки» [100] – системи, внутрішня логіка яких залишається прихованою від користувача. У зв'язку з цим виникає потреба в розвитку галузі пояснюваного штучного інтелекту (XAI) [101], яка прагне подолати розрив між ефективністю та зрозумілістю моделей.

Концептуально, XAI [102] – це сукупність методів, які дозволяють робити інтерпретації рішень моделей машинного навчання доступними для розуміння людини, водночас зберігаючи або лише мінімально жертвуючи продуктивністю. Йдеться не лише про розкриття внутрішньої структури моделей, а й про створення семантичного моста між формальними обчисленнями та людським сприйняттям. Ця сфера охоплює теоретичні дослідження, інженерні рішення, психологічні аспекти сприйняття інтерпретацій, а також нормативно-правові питання, пов'язані з відповідальністю та довірою [103, 104].

Науково обґрунтована потреба у XAI виникає з кількох джерел. По-перше, це вимога регуляторної підзвітності – у багатьох юрисдикціях ухвалені закони, які вимагають пояснення автоматизованих рішень, особливо в контексті захисту прав користувачів [105]. По-друге, це етична проблема: без пояснення причин автоматичних дій система втрачає елемент справедливості, адже користувач не може оскаржити або навіть зрозуміти отримане рішення. По-третє, з інженерної точки зору, відсутність пояснюваності унеможливорює виявлення помилок, зниження ризиків та валідацію моделей [106], що критично важливо в галузях із високою ціною помилки. По-четверте, наявність пояснень дозволяє адаптувати моделі до змін середовища, виконувати цільове налагодження (debugging) та виявляти упередження.

Теоретична основа XAI лежить у понятті каузальності (причинності) – ідеї про те, що хороше пояснення має відповідати причинному ланцюгу подій або змін [107]. У практиці XAI це означає, що модель повинна не лише

виявити кореляції між ознаками та мітками, але й бути здатною вказати, які саме ознаки були визначальними, і як зміна цих ознак могла би вплинути на результат. При цьому важливим аспектом є локальність пояснень: багато підходів до ХАІ зосереджуються на поясненні конкретного рішення, а не поведінки моделі в цілому.

Існує також фундаментальне розрізнення між двома категоріями ХАІ: інтрузивні методи, які вбудовані в саму модель або створюються паралельно з нею (наприклад, інтерпретовані моделі на зразок дерева рішень [108] або генеративні контрприкладі), та постхокові методи (post-hoc) [109], які застосовуються вже після навчання моделі для пояснення її дій, наприклад, через візуалізацію важливості ознак, локальні апроксимації [110] або обчислення вкладень.

Найбільш відомі підходи до пояснюваності включають локальні інтерпретатори (наприклад, LIME [111, 112], SHAP [113, 114]), методи атрибуції важливості (наприклад, Grad-CAM [115], Integrated Gradients [116]), спрощені моделі-замінники (surrogate models), контрфактичні пояснення (які демонструють, як слід було би змінити вхідні дані, щоби змінити рішення моделі), а також природномовні генератори пояснень, які комбінують неймережеві методи з лінгвістичною інтерпретацією.

Одним із найефективніших напрямів є використання локальних інтерпретаторів, які зосереджуються на поясненні окремих прогнозів моделі. Замість того, щоб аналізувати всю модель цілком, ці методи намагаються відповісти на питання, чому модель ухвалила конкретне рішення для конкретного прикладу. До таких підходів належить, зокрема, Local Interpretable Model-Agnostic Explanations (LIME), що базується на апроксимації локального простору навколо передбачення за допомогою простої інтерпретованої моделі, наприклад, лінійної регресії або дерева рішень. Штучно згенеровані сусідні приклади використовуються для навчання цієї локальної моделі, яка дозволяє оцінити, які саме ознаки найбільше вплинули на рішення моделі в цій точці.

Схожим за ідеологією, але потужнішим у плані стабільності інтерпретацій, є метод SHAP (SHapley Additive exPlanations), що базується на концепції Шеплі з теорії кооперативних ігор. Цей підхід дозволяє розраховувати внесок кожної ознаки в кінцевий результат, розглядаючи всі можливі комбінації ознак і усереднюючи їхню маржинальну важливість. SHAP забезпечує як локальну, так і глобальну інтерпретацію, що робить його особливо корисним для моделей з великою кількістю параметрів.

Атрибуційні методи [117], на відміну від моделей-агностиків, тісно інтегруються у структуру самої нейромережі. Вони аналізують градієнти або активації нейронів для виявлення важливих фрагментів вхідних даних. Так, метод Integrated Gradients [118] передбачає поступове зміщення вхідного вектора від базового (наприклад, нульового) до цільового і накопичення градієнтів уздовж цього шляху. Це дозволяє обчислити інтегральну оцінку впливу кожного елемента вхідного вектора на остаточне рішення. У свою чергу, Grad-CAM (Gradient-weighted Class Activation Mapping), орієнтований переважно на візуальні дані, використовує градієнти, що проходять через останні згорткові шари нейромережі, для генерації теплових карт, які вказують, які ділянки зображення (або абстрактно – ознаки) стали вирішальними для класифікації.

У ще одному класі підходів – контрфактичних поясненнях – головна увага зосереджена не на тому, чому модель ухвалила певне рішення, а на тому, що потрібно було б змінити у вхідних даних, аби рішення стало іншим. Контрфактичне пояснення формулюється як мінімально можлива модифікація прикладу, яка змінює вихід моделі. Цей метод особливо цінний у контексті персоналізованих рішень, наприклад, коли йдеться про психоемоційну підтримку: контрфактичний сценарій може вказати, які саме фактори треба модифікувати, аби зменшити ймовірність ризикової поведінки.

Усе більшого поширення також набувають підходи, засновані на генерації пояснень у природній мові [119]. Такі моделі використовують трансформери або інші архітектури зі здатністю до умовної генерації, які навчаються на парах «вхід – пояснення» і здатні сформулювати логічне та

послідовне тлумачення рішення моделі у зрозумілій людині формі. Цей підхід дозволяє подолати розрив між математичними обґрунтуваннями рішень і потребою користувача в когнітивно доступному поясненні. Вони також відкривають нові горизонти для етичного та прозорого впровадження систем ШІ у чутливі галузі, зокрема в системи моніторингу психоемоційного стану населення.

Складність побудови пояснюваного ШІ полягає не лише у високому рівні абстракції моделей, а й у множинності цільових аудиторій пояснення. Для розробника важливі технічні пояснення, які дозволяють вдосконалювати модель; для користувача – прості, лаконічні та когнітивно зрозумілі відповіді; для регулятора – формально достовірні й перевірювані трактування. Ця багатовекторність накладає вимогу до ХАІ бути адаптивним, тобто гнучко реагувати на контекст використання.

Ще один ключовий момент полягає у співвідношенні пояснюваності та продуктивності. Високоточні моделі часто стають складнішими, менш прозорими; водночас прості моделі легше пояснити, але вони менш адаптовані до складних патернів даних. Це породжує так звану діалектику ХАІ: компроміс між складністю і пояснюваністю, який намагаються подолати шляхом створення гібридних систем або метамоделей, що діють як «коментатори» основної моделі.

Сучасні дослідження у сфері ХАІ також виходять за межі класичної інтерпретованості і включають такі питання, як довіра до моделі, когнітивне навантаження користувача, неперервність пояснень у разі оновлення моделі, а також метрики оцінки якості пояснень (наприклад, стабільність, правдивість, узгодженість та корисність пояснень). Крім того, особливу увагу приділяють оцінці пояснень не лише як технічних артефактів, але і як елементів соціальної взаємодії між системою та користувачем, що формують етику взаємодії з алгоритмами.

Таким чином, пояснюваний штучний інтелект є не лише технічним інструментом, а й філософською, когнітивною та етичною парадигмою, що переосмислює відносини між людиною і системою автоматизованого прийняття

рішень. Його розвиток є фундаментальним кроком до побудови надійного, відповідального та людиноорієнтованого штучного інтелекту. У майбутньому саме ХАІ може стати ключовим чинником прийняття рішень у розподілених системах, автономних агентів та персоналізованих інтелектуальних помічників, де прозорість і довіра є так само важливими, як і точність.

У дисертаційній роботі з наведених підходів використано пост-хок методи пояснення, сумісні з трансформерними моделями: атрибуцію внеску окремих токенів (Transformers Interpret) і візуалізацію ваг механізму уваги (BertViz) для методу аналізу прояву ПТСР (п. 4.2). Для методу розширення множини ознак пояснюваність забезпечується конструктивно: індекс ризику є адитивним і декомпозується на внески базової частки ПТСР-повідомлень, коморбідності та негативного емоційного контексту (п. 2.5). Методи LIME, SHAP, Grad-CAM і контрфактичні пояснення в роботі не застосовуються та наведені лише для повноти огляду.

1.7. Мета та завдання дослідження

Проведений аналіз (п. 1.1–1.6) виявив три невирішені проблеми: відсутність контрольованого відмежування ознак ПТСР від тематично суміжних станів на рівні дописів; орієнтацію відомих моделей на один домінуючий розлад; оцінювання поведінкових ризиків без урахування їх часової динаміки. Об'єкт і предмет дослідження визначено у Вступі; з наведених проблем випливають мета і завдання дослідження.

Метою дослідження є підвищення точності виявлення ознак посттравматичного стресового розладу і супутніх психічних розладів та розширення можливостей оцінювання ризиків соціально небезпечної поведінки громадян за інтелектуальним аналізом користувацького контенту засобами контекстно-орієнтованих нейромережевих моделей і розширення простору ознак.

Під точністю виявлення розуміються показники класифікації на рівні повідомлення: Accuracy, Precision, Recall, F1-score. Розширення можливостей

оцінювання ризиків полягає в переході від рішень щодо окремих повідомлень до скринінгового часового індексу ризику на рівні користувача, який інтегрує оцінки прояву ПТСР, інших психічних розладів та емоційного контексту повідомлень. Супутніми названо психічні розлади, ознаки яких можуть виявлятися в текстах одночасно з ознаками ПТСР; їх зв'язок саме з ПТСР на використаних наборах даних не перевірявся.

Для досягнення поставленої мети ставляться такі завдання:

1. Провести аналіз сучасних методів, моделей і засобів автоматизованого виявлення посттравматичного стресового розладу та інших психічних розладів за текстовим користувацьким контентом.

2. Удосконалити метод аналізу прояву посттравматичного стресового розладу в користувацькому контенті з урахуванням контекстних залежностей, специфічних для ПТСР.

3. Удосконалити метод виявлення психічних розладів за аналізом користувацького контенту з можливістю ідентифікації кількох розладів одночасно.

4. Розробити метод розширення множини ознак користувацького контенту для попередження соціально небезпечної поведінки громадян із ПТСР.

5. Розробити програмну систему інтелектуального аналізу користувацького контенту та експериментально дослідити ефективність запропонованих методів.

Структурно завдання дослідження співвіднесено з трьома науковими результатами таким чином: завдання 2 забезпечує метод аналізу прояву ПТСР, включно з формуванням спеціалізованого навчального набору даних; завдання 3 відповідає методу виявлення психічних розладів; завдання 4 – методу розширення множини ознак для попередження соціально небезпечної поведінки; завдання 5 забезпечує їх програмну реалізацію та експериментальну перевірку. Таке співвіднесення використано далі лише для організації викладу і не змінює назв методів, сформульованих у положеннях наукової новизни.

РОЗДІЛ 2. ТЕОРЕТИЧНІ ОСНОВИ ТА ФОРМАЛІЗАЦІЯ МЕТОДІВ АНАЛІЗУ ПСИХІЧНОГО СТАНУ ЗА КОРИСТУВАЦЬКИМ КОНТЕНТОМ

2.1. Формалізація предметної області та загальної постановки задачі

Загальна ідея роботи полягає у розробленні підходу до аналізу психічного стану громадян із ПТСР шляхом інтелектуальної обробки користувацького контенту.

Запропонований підхід дозволяє не лише визначити наявність ознак ПТСР, але й оцінити ризик соціально небезпечної поведінки, пов'язаної з психічними розладами та наведений на рис. 2.1.

У роботі аналіз виконується для текстового користувацького контенту, пов'язаного з конкретним користувачем i , для задачі оцінювання поведінкових ризиків, з часовими мітками повідомлень. Нехай $U = \{u^1, \dots, u_G\}$ – множина користувачів, а профіль користувача u містить N_u повідомлень $m_{u,i}$ з часовими мітками $t_{u,i} \in \mathbb{R}_+$ (момент публікації повідомлення, виміряний у добах від початку спостереження; повідомлення впорядковано за зростанням $t_{u,i}$).

$$P_u = \{m_{u,i}, t_{u,i}\}_{i=1}^{N_u}. \quad (2.1)$$

Для застосування нейромережових моделей повідомлення відображається у числовий простір ознак $X \subseteq \mathbb{R}^V$. Відображення φ узагальнює токенізацію та формування контекстного представлення тексту засобами трансформерних моделей із розділеним механізмом уваги [120, 121]; конкретні способи його реалізації наведено у розділі 3 та опубліковано в авторських працях [158, 160].

$$\varphi: \mathcal{T} \rightarrow X, \quad x = \varphi(m), \quad (2.2)$$

де $\mathcal{T}$ – множина текстових повідомлень, кожне з яких є скінченною впорядкованою послідовністю символів алфавіту $\mathcal{A}$, тобто $\mathcal{T} \subseteq \mathcal{A}^*$; $m \in \mathcal{T}$ – окреме повідомлення.

У межах дисертації розв'язуються три взаємопов'язані задачі: оцінювання прояву ПТСР, виявлення кількох психічних розладів та розширення множини ознак для оцінювання ризику соціально небезпечної поведінки. Для компактності математичного запису оператор, що реалізує метод аналізу прояву ПТСР у користувацькому контенті далі позначено T_1 , метод виявлення психічних розладів за аналізом користувацького контенту – T_2 , а метод розширення множини ознак користувацького контенту для попередження соціально небезпечної поведінки громадян із ПТСР – T_3 . Функціональну взаємодію методів наведено на рис. 2.1. Отже, T_1 , T_2 і T_3 є операторами (відображеннями), які реалізують відповідно методи 1–3 (М1–М3 на рис. 2.1); сигнатури операторів наведено в п. 2.2–2.5.

Множини класів визначено так: для методу аналізу прояву ПТСР – $Y_1 = \{0, 1\}$ (відсутність або наявність ознак ПТСР); для методу виявлення психічних розладів – множина розладів $D = \{d^1, \dots, d_K\}$, $K = 5$, кожному з яких відповідає незалежна бінарна мітка $y_k \in \{0, 1\}$; для методу розширення множини ознак результатом є не клас, а числовий індекс ризику $R_{u,t} \in [0, 1]$, визначений на рівні користувача.

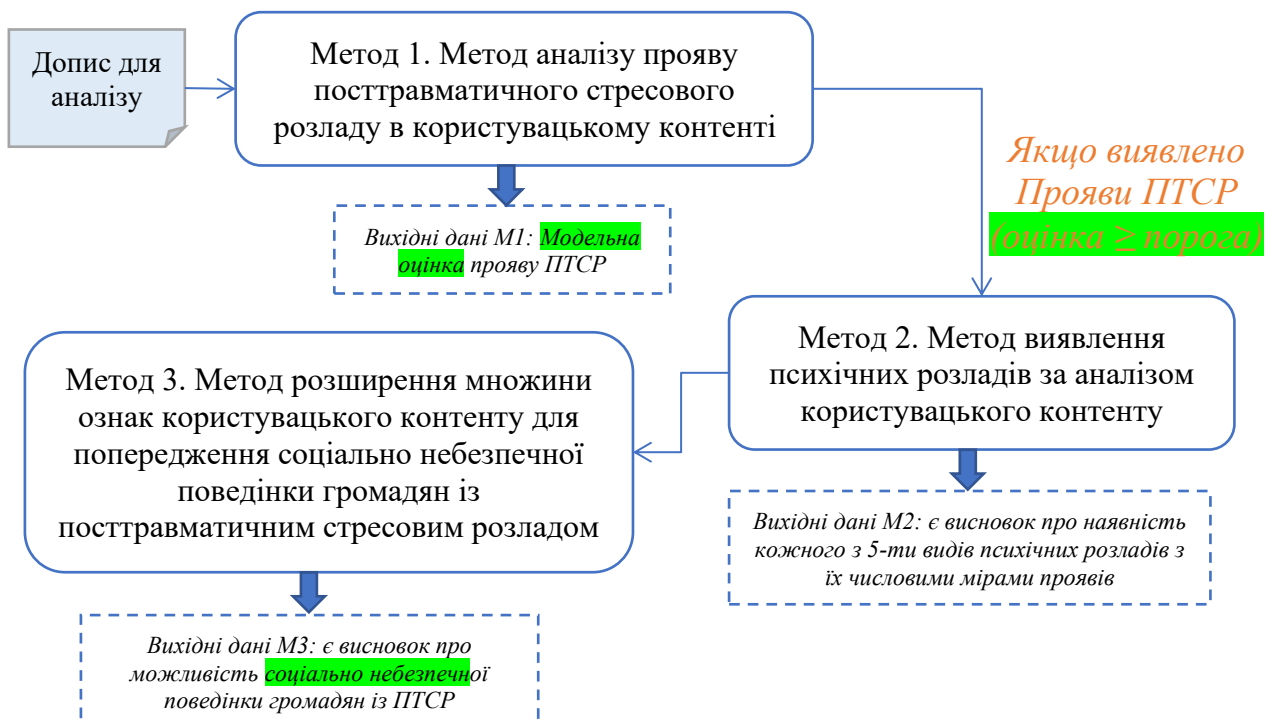


Рис. 2.1 – Підхід до аналізу психічного стану громадян із ПТСР

Ця мета досягається шляхом узгодженого застосування трьох методів у повному або умовно-послідовному режимі (п. 2.5); на рис. 2.1 показано умовно-послідовний режим, у якому текст допису надходить на вхід кожного з методів, а зв'язок між методами 1 і 2 позначає лише умову активації. Перший метод призначений для оцінки прояву ПТСР у користувацькому контенті на основі виявлення характерних ознак та їх кількісної інтерпретації відносно встановленого порогу.

У разі перевищення порогового значення активується другий метод, спрямований на глибший аналіз тексту з метою виявлення наявності інших психічних розладів, їх типів та кількісних характеристик прояву.

Третій метод забезпечує розширення множини ознак користувацького контенту для скринінгового оцінювання ризику соціально небезпечної поведінки осіб із виявленими ознаками ПТСР шляхом використання інтегрованих висновків попередніх етапів.

Вихідними даними є модельна оцінка прояву ПТСР, характеристика виявлених психічних розладів за видами та мірами їх прояву, а також скринінговий індекс ризику соціально небезпечної поведінки громадян, що мають ознаки ПТСР.

Таким чином, розроблений підхід до аналізу психічного стану громадян із ПТСР забезпечує інтелектуальне виявлення психологічних ризиків шляхом автоматизованої обробки користувацького контенту. Запропонована послідовність методів дозволяє не лише встановлювати наявність ознак ПТСР, але й здійснювати диференційовану оцінку супутніх психічних розладів, що дає змогу отримати комплексне уявлення про психоемоційний стан особи. Інтеграція результатів попередніх етапів у скринінговий індекс ризику дає фахівцеві підставу звернути увагу на користувачів зі стійко підвищеним рівнем ризику; остаточне рішення ухвалює людина (п. 3.8). Такий підхід має потенціал для практичного

застосування в системах моніторингу психічного здоров'я та сприяє розвитку превентивних механізмів у сфері громадської безпеки.

Взаємодія методів має умовно-послідовний характер. Метод 1 (оператор T_1) формує кількісну оцінку прояву ПТСР; якщо вона перевищує встановлений поріг, *активується метод 2 (оператор T_2)* для виявлення супутніх психічних розладів. При цьому результат T_1 є умовою активації T_2 , а не його вхідним вектором ознак: T_2 аналізує сформоване за (2.2) представлення x того самого текстового повідомлення. Метод 3 (оператор T_3) використовує результати попередніх етапів разом з емоційно-семантичними та поведінковими характеристиками під час формування розширеної множини ознак для оцінювання ризику соціально небезпечної поведінки.

2.2. Метод аналізу прояву ПТСР у користувацькому контенті

Для повідомлення m з представленням $x = \varphi(m)$ метод формує значення p_{PTSD} , що інтерпретується як модельна оцінка належності тексту до класу з ознаками ПТСР. У реалізованій конфігурації T_1 параметризується контекстно-орієнтованою трансформерною моделлю. Сигнатура оператора: $T_1: X \rightarrow [0, 1]$.

$$p_{PTSD} = T^1(x; \theta^1), \quad p_{PTSD} \in [0, 1]. \quad (2.3)$$

Класифікаційне рішення отримується порогованням модельної оцінки. Поріг τ_{PTSD} є параметром процедури прийняття рішення і не змінює безперервного виходу моделі.

$$\hat{y}_{PTSD} = \mathbb{1}[p_{PTSD} \geq \tau_{PTSD}]. \quad (2.4)$$

Ознакове представлення тексту доцільно розглядати як поєднання лексичної, контекстної та позиційної складових $x = (x^{\text{lex}}, x^{\text{ctx}}, x^{\text{pos}})$. У фактичній реалізації ключовими є контекстна та позиційна складові DeBERTa: роздільне моделювання змісту й позиції та відносне позиційне зміщення дають змогу враховувати залежності між словами і їх положенням у тексті [120, 121].

Окремим компонентом методу є формування спеціалізованої навчальної вибірки: негативний клас містить не лише тексти без ознак психічних розладів, а й тексти з проявами інших розладів. Це сприяє процесу відмежування ПТСР від семантично близьких станів; джерела та правила формування відповідної вибірки наведено у п. 3.5.1, а принцип помилко-орієнтованого формування такої вибірки опубліковано в авторських працях [158, 160].

Ефект щодо покращення аналізу проявів ПТСР за запропонованим підходом зображено на рис. 2.2.

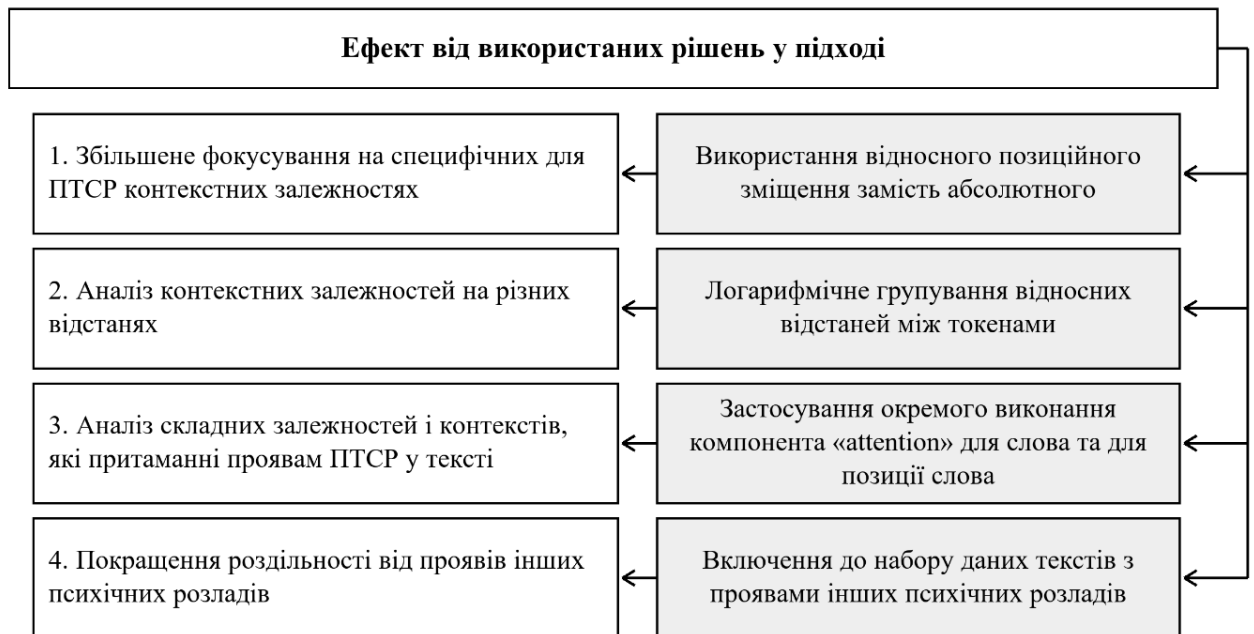


Рис. 2.2 – Ефект щодо покращення аналізу проявів ПТСР за запропонованим підходом

Зокрема, шляхом використання відносного позиційного зміщення замість абсолютного вдалося досягти збільшеного фокусування на специфічних для ПТСР контекстних залежностях, шляхом логарифмічного групування відносних відстаней між токенами (параметр `position_buckets` у конфігурації моделі [140]) вдалося проводити аналіз контекстних залежностей на різних відстанях, шляхом застосування окремого виконання компоненту «attention» для слова та для позиції слова вдалося виконувати

аналіз складних залежностей і контекстів, які притаманні проявам ПТСР у тексті, шляхом включення до набору даних текстів з проявами інших психічних розладів вдалось досягти покращення роздільності від проявів інших психічних розладів. Ефекти 1–3 є наслідком властивостей використаної контекстно-орієнтованої моделі, тоді як ефект 4 забезпечується запропонованим способом формування навчального набору даних; відмінність методу полягає саме в поєднанні цих складових для задачі виявлення ПТСР.

На відміну від проаналізованих методів автоматизованого виявлення ПТСР за текстовим контентом [17–23], запропонований метод орієнтований не лише на виявлення загальних мовних маркерів психічного стану, а на виділення контекстних залежностей, специфічних для ПТСР. Це забезпечується поєднанням контекстно-позиційного представлення тексту [120, 121] зі спеціалізованим формуванням нецільового класу, до якого включено тексти з проявами інших психічних розладів [158, 160], що дозволяє підвищити роздільність між ПТСР та семантично близькими психічними станами.

2.3. Метод виявлення психічних розладів за аналізом користувацького контенту

Задача, що розглядається, відноситься до задач класифікації. Можна виокремити три типи класифікації: бінарну; з кількома класами (багатокласову) та з кількома мітками (багатоміткову). Бінарна класифікація відносить об'єкт до одного із двох взаємовиключних класів. Багатокласова передбачає, що кількість класів більше двох, але при цьому, класи є взаємовиключними, тобто одночасно об'єкт може відноситися лише до одного з класів. Ці два типи класифікації часто позначають як класифікацією з однією міткою (single-label classification – SLC), оскільки в ній, при навчанні класифікатора, використовується лише одна мітка класу.

Класифікація з багатьма мітками (multi-label classification - MLC) відносить об'єкт до кількох класів одночасно, які при цьому не є взаємовиключними.

Можна розглянути MLC як окремий випадок SLC. Для цього розглянемо загальну постановку завдання класифікації. Нехай задано множину об'єктів, що класифікуються $X = \{x_1, x_2, \dots, x_B\}$ і множину міток класів $Y = \{y_1, y_2, \dots, y_C\}$. При цьому існує невідоме відображення множини об'єктів у множину класів (класифікатор): $h : X \rightarrow Y$, значення яких відомі на кінцевій навчальній вибірці (розміром Q): $X^Q = \{(x_1, y_1), (x_2, y_2), \dots, (x_Q, y_Q)\}$.

Отже, потрібно побудувати класифікатор $h : X \rightarrow Y$, який кожному об'єкту $x_i \in X$ ставить у відповідність $y_i \in Y$. При цьому, якщо $C = 2$, то маємо бінарну класифікацію, а при $C > 2$ – багатокласову.

Для MLC, навчальний набір якого міститиме декілька полів міток класів, потрібно вже ансамбль класифікаторів: $H_x = \{h_1, h_2, \dots, h_C\}$, розмір якого дорівнює кількості міток класу.

Виділяють два основні підходи до побудови MLC-моделей класифікації: трансформації даних та адаптації алгоритмів. Перший полягає у перетворенні навчального набору даних. Другий передбачає адаптацію звичайних алгоритмів машинного навчання для побудови MLC-класифікатора.

Відомі наступні методи трансформації даних: бінарної релевантності (binary relevance, BR), ланцюги класифікаторів (classifier chain, CC) та набір потужності міток (label-powerset, LP).

Метод BR. В основі методу лежить перетворення задачі MLC до кількох завдань бінарної класифікації з однією міткою. При цьому для кожного класу, якому може належати об'єкт, у навчальному наборі створюється окреме поле, де, для прикладу, ставиться 1, якщо він належить класу, і 0 – у іншому випадку. Зауважимо, що метод BR має два основні недоліки. Перший – необхідність навчати множину класифікаторів, що може

призвести до значних обчислювальних витрат. Другий – він не враховує залежність міток. Проте, щоб вирішити проблему залежності міток, можна використовувати інший підхід: ланцюги класифікаторів, тобто вибудувати бінарні класифікатори в послідовний ланцюг, де входами класифікаторів є не тільки об'єкти, а й виходи попереднього класифікатора в ланцюзі.

Таким чином, класифікаційні задачі, залежно від кількості класів та способу віднесення об'єктів до них, можуть реалізовуватись у різних постановках: бінарній, багатокласовій або багатоміткової. У випадку багатоміткової класифікації постає необхідність у побудові кількох класифікаційних моделей, що зумовлює як зростання обчислювальної складності, так і потребу врахування взаємозв'язків між мітками. Розв'язання задач такого типу можливе завдяки підходам, що ґрунтуються на трансформації даних або адаптації алгоритмів, зокрема методам бінарної релевантності та ланцюгів класифікаторів. Кожен із підходів має свої переваги та обмеження, вибір яких визначається специфікою задачі та вимогами до моделі.

Мета методу виявлення психічних розладів – одночасне виявлення кількох психічних розладів у повідомленні без втрати менш виражених, але клінічно значущих ознак.

Ідея запропонованого підходу (рис. 2.3) полягає у врахуванні протиріччя, яке виникає при багатокласовій класифікації, а саме виявлення найбільш вираженого психічного розладу та втрата, при цьому, інших, менш вагомих ознак, які можуть вказувати на менш виражений але більш суттєвий психічний розлад. Наявність такого протиріччя призводить до втрати часу та упущання можливості для надання психологічної допомоги при початкових проявах психічних розладів [124].

На противагу багатокласовій класифікації, пропонується використати багатоміткову класифікацію. Перевагою такої класифікації є те, що MLC враховує випадки, коли характерні ознаки можуть відноситися до більш ніж одного класу одночасно та при цьому класи не будуть взаємовиключними [125].

Зазначимо, що запропоноване дозволить уникнути проблеми втрати менш яскраво виражених ознак розладів, однак залишає проблему низької точності класифікації, притаманної як багатокласовій так і багатоміткової класифікаціях. Проблема низької точності пов'язана з відсутністю коректно сформованих навчальних даних, які б були розмічені фахівцями, та при цьому містили б достатню кількість зразків для навчання [126].

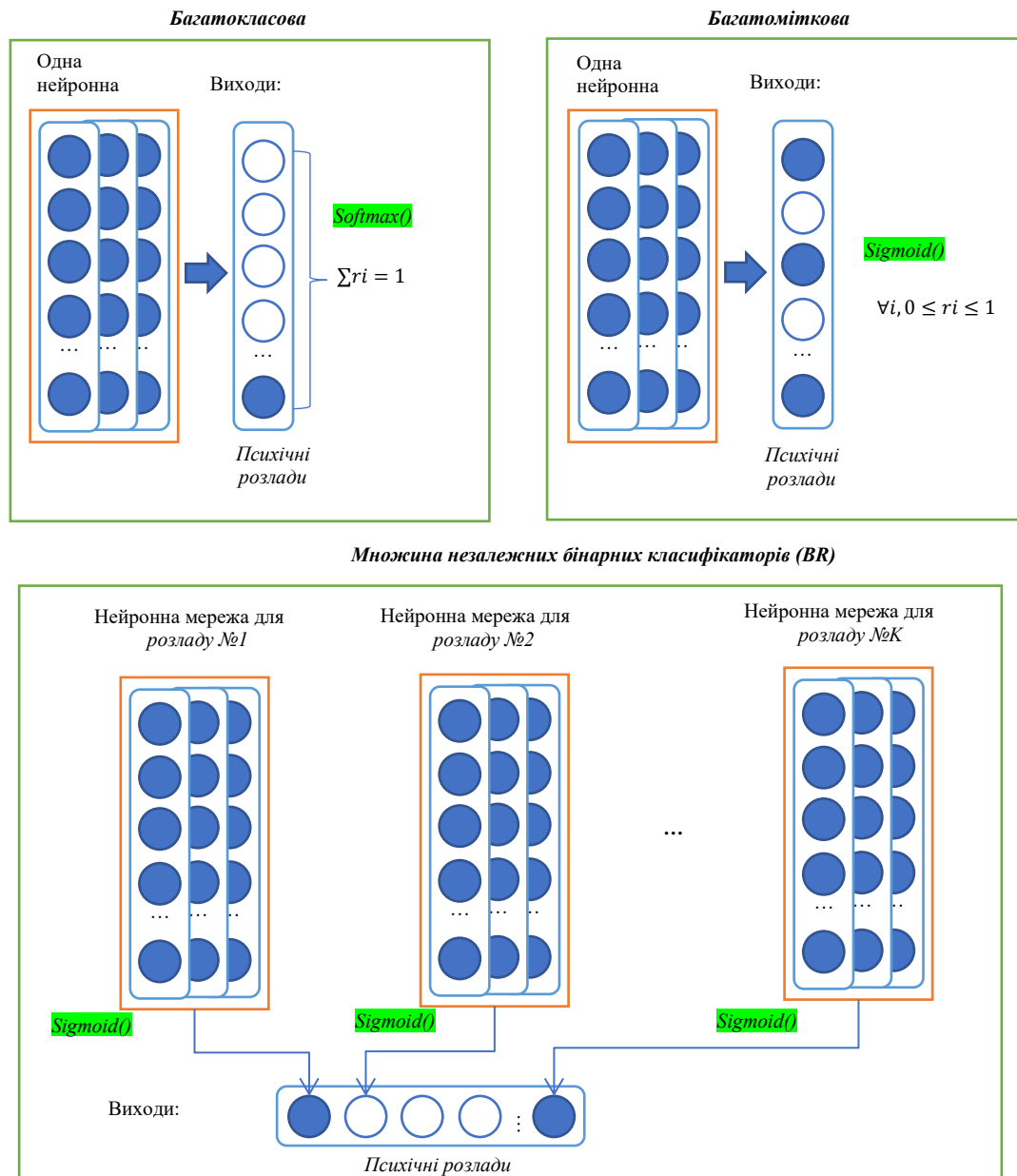


Рис. 2.3 – Підходи до класифікації психічних розладів

Іншою проблемою є те, що навіть при наявності розмітки, деякі види психічних розладів мають схожі ознаки в текстовому контенті. Висувається

припущення, що для підвищення показників коректної ідентифікації психічних розладів необхідно коректно формувати навчальні вибірки, які б містили цільовий клас психічного розладу, та інший, що є сукупністю інших розладів та текстами що не містять психічних розладів. Наведене пропонується реалізувати шляхом використання MLC-моделей класифікації з трансформацією даних (метод BR). Тобто, пропонується декомпонувати задачу MLC до кількох завдань бінарної класифікації з однією міткою.

Вибір декомпозиції на бінарні класифікатори замість єдиної багатоміткової моделі обґрунтовано трьома чинниками. По-перше, доступні набори даних розмічено за принципом однієї мітки на текст, тому спільна MLC-модель навчається на векторах міток, у яких супутні розлади не позначені, і сприймає їх як негативні приклади; бінарна модель кожного розладу цього припущення не потребує. По-друге, для кожного розладу формується власна навчальна вибірка, у нецільовому класі якої пропорційно представлені інші розлади, що дає змогу моделі вивчати саме розрізнявальні ознаки цього розладу. По-третє, індивідуальні пороги за індексом Юдена дозволяють незалежно налаштувати чутливість кожної моделі. Агрегація результатів здійснюється формуванням вектора рішень без взаємовиключення класів (2.14); експериментальне порівняння з багатокласовою та багатомітковою моделями наведено в п. 4.3 (табл. 4.5).

Далі наведено модель процесу та основні кроки запропонованого підходу. Модель використовує наступні позначення: m – текст (повідомлення) для аналізу; розглядається аналіз користувацького контенту у соціальних мережах і, відповідно, є обмеження на його довжину – кількість токенів тексту не повинна перевищувати 128 токенів; f_1 – токенізатор WordPiece [127], який реалізує відображення f_1 (2.5):

$$f_1: m \rightarrow tkn, \quad (2.5)$$

де tkn токенізований текст m у вигляді упорядкованої послідовності токенів; D – множина психічних розладів (mental disorder), які ідентифікуються за tkn :

$$D = \{d_1, d_2, \dots, d_K\}, d_i = (id_{DSO,i}, Name_{DSO,i}), i = 1, 2, \dots, K, \quad (2.6)$$

де K – кількість психічних розладів, id_{DSO} – ід ментального розладу (при $id_{DSO} = 0$ психічний розлад відсутній), $Name_{DSO}$ – назва психічного розладу;

DS – набір даних з розміченими, за приналежністю до певного психічного розладу d_i , текстами, за яким будуть створюватись набори даних для тонкого налаштування моделей глибокого навчання:

$$DS = \{DS_1, DS_2, \dots, DS_S\}, DS_i = (m_i, id_{DSO,i}), i = 1, 2, \dots, S, \quad (2.7)$$

де S – кількість зразків у DS ; Df – множина наборів даних для тонкого налаштування моделей глибокого навчання для бінарної класифікації; кожен набір даних із Df містить розмічені токенизовані за (2.5) tkn для конкретного психічного розладу ($Cls = id_{DSO}$), або інших розладів ($Cls = 0$); Df отримується відображенням f_2 :

$$f_2: DS \rightarrow Df, \quad (2.8)$$

де Df :

$$Df = \{Df_1, Df_2, \dots, Df_K\}, Df_i = \{Df_{i1}, Df_{i2}, \dots, Df_{iS_i}\}, i = 1, 2, \dots, K \quad (2.9)$$

$$Df_{ij} = (tkn_{ij}, Cls_{ij}), j = 1, 2, \dots, S_i,$$

де S_i – кількість зразків у наборі даних Df_i ; зазначені набори даних Df_i формуються для бінарної класифікації, тобто мають цільовий клас зі зразками, які відповідають конкретному розладу $d_i \in D$ та нецільовий клас (всі інші розлади або їх відсутність); для запобігання плутанини психічних розладів між собою та враховуючи, що в одному тексті можуть бути і прояви інших розладів з множини D , у нецільовій категорії записи формуються за певними правилами [128]: (а) кількість зразків у нецільовій групі відповідає, або наближається до кількості зразків у цільовій; (б) нецільова група складається із рівних пропорцій решти зразків із DS з іншими видами психічних розладів та даних без проявів психічних розладів.

M – множина моделей глибокого навчання, отримана за допомогою відображення f_3 , яке реалізує тонке налаштування DLM за допомогою Df :

$$f_3: (DLM, Df) \rightarrow M, \quad (2.10)$$

де DLM – модель глибокого навчання DistilBERT, M – множина тонко налаштованих моделей глибокого навчання для кожного психічного розладу $d_i \in D$:

$$M = \{\mu_1, \mu_2, \dots, \mu_K\}, \mu_i = (mod_i, \tau_i), i = 1, 2, \dots, K, \quad (2.11)$$

де mod – тонко налаштована модель глибокого навчання DLM за допомогою набору даних $Df_i \in Df$ для конкретного психічного розладу $d_i \in D$,

τ – поріг для визначення психічного розладу d_i , який визначається за індексом Юдена – значенням точки на ROC-кривій, яка відповідає найкращому співвідношенню чутливості та специфічності [129].

r – результат застосування $\mu_i \in M, i = 1, 2, \dots, K$ до довільного токенозованого тексту tkn :

$$r = \{r_1, r_2, \dots, r_K\}, r_i = (p_i), i = 1, 2, \dots, K, \quad (2.12)$$

де p_i – числове значення впливу розладу d_i для токенозованого тексту tkn ; $r_i \in [0,1]$ яке вказує на величину прояву психічного розладу $d_i \in D$. Значення p_i є вихідним значенням останнього шару моделі mod_i , отримане сигмоїдною функцією активації. r отримується за допомогою відображення f_4 :

$$f_4: tkn \rightarrow r, \quad (2.13)$$

де r містить числові значення проявів розладів для кожного $d_i \in D$; аналіз наявності чи відсутності психічного розладу визначається, порівнянням отриманого числового значення p_i з пороговим значенням τ .

Важливо, що застосування бінарної релевантності означає незалежне навчання окремих класифікаторів, але не вводить припущення про статистичну або клінічну незалежність самих психічних розладів. Навпаки, неврахування залежностей між мітками є відомим обмеженням BR, що враховано при інтерпретації результатів [125, 126].

Для кожного розладу d_k використовується власний поріг τ_k . У варіанті з адаптивним порогуванням його визначають на ROC-кривій за індексом Юдена [129]:

$$\tau_k^* = \arg \max_{\tau} \{TPR_k(\tau) - FPR_k(\tau)\}. \quad (2.14)$$

Тоді вектор рішення для K розладів має вигляд $\hat{y} = (\mathbb{1}[p_1 \geq \tau_1^*], \dots, \mathbb{1}[p_K \geq \tau_K^*])$. Це архітектурно допускає одночасне виявлення кількох станів без примусового вибору одного домінуючого класу. Сигнатура оператора: $T_2: X \rightarrow [0, 1]^K$ з подальшим пороговуванням до вектора рішень із $\{0, 1\}^K$.

2.4. Метод розширення множини ознак користувацького контенту для попередження соціально небезпечної поведінки громадян із посттравматичним стресовим розладом

Для попередження соціально небезпечної поведінки громадян із ПТСР необхідно навести розширену множину ознак. Такими ознаками в межах дослідження є: загальна тональність тексту, тональність за реченнями, кількість смайлів, кількість граматичних помилок, рівень агресії, емоції за NRC Emotion Lexicon, різка зміна тональності між реченнями, наявність погроз/небезпечної лексики, позитивні емоції в NRC, кількість негативних речень. Опис ознак з прикладами наведено в таблиці 2.1.

Основною метою даного підходу є побудова комплексної ознакової моделі, здатної не лише виявити ознаки емоційної нестабільності, агресії чи суїцидальних схильностей, але й оцінити динаміку емоційного фону повідомлень, їхню внутрішню когерентність та наявність лінгвістичних маркерів критичних станів. У межах дослідження було запропоновано метод розширення ознакового простору, що базується на поєднанні класичних емоційних індикаторів (зокрема, за допомогою NRC Emotion Lexicon) із додатковими синтаксико-семантичними та прагматичними характеристиками тексту.

У сучасних умовах зростання соціальної напруги, обумовленої як глобальними викликами, так і локальними кризовими ситуаціями, питання виявлення потенційно небезпечної поведінки громадян, що перебувають у стані ПТСР, набуває особливої актуальності.

Таблиця 2.1

Множина ознак користувацького контенту для попередження соціально небезпечної поведінки громадян

№	Ознака аналізу	Опис	Тип значення	Приклади/Деталі
1	Загальна тональність тексту	Оцінка всього тексту на позитив/негатив/нейтраль	позитив/негатив/нейтраль	Використовується для фільтрації загального емоційного фону
2	Тональність за реченнями	Оцінка кожного речення окремо	список тональностей	['негатив', 'нейтраль', 'негатив', 'позитив']
3	Кількість смайлів	Загальна кількість смайлів у тексті	ціле число	Використовується як індикатор емоційного вираження
4	Кількість граматичних помилок	Кількість помилок у синтаксисі, пунктуації або орфографії	ціле число	0 – без помилок, >3 – може вказувати на емоційну нестабільність
5	Рівень агресії	Визначає рівень агресії у тексті (на основі словника або моделі)	float (0,0 - 1,0)	>0,6 – висока агресія; 0,3–0,6 – середня
6	Емоції за NRC Emotion Lexicon	Визначення наявності емоцій у тексті (joy, anger, fear, sadness, trust тощо)	набір емоцій	['anger', 'fear', 'disgust'], ['joy', 'trust']
7	Різка зміна тональності між реченнями	Чи є перехід між емоційно протилежними реченнями	так/ні	Напр., позитив → негатив
8	Наявність погроз/небезпечної лексики	Слова чи фрази, які вказують на загрозу або небезпечні наміри	так/ні	Словник небезпечних слів: “вб’ю”, “ненавиджу”, “знищу”, “розіб’ю”
9	Позитивні емоції в NRC	Чи є серед емоцій NRC хоча б одна позитивна (joy, trust, anticipation)	так/ні	Використовується для виявлення апатії або повної відсутності позитиву
10	Кількість негативних речень	Скільки з речень мають негативну тональність	ціле число	Якщо >2 – можливе накопичення негативу

Одним з ефективних інструментів у цьому контексті є аналіз користувацького контенту у відкритих інформаційних середовищах – соціальних мережах, форумах, блогах тощо. Враховуючи складність психологічного стану осіб з ПТСР, критично важливо застосовувати розширений підхід до побудови множини ознак, які відображають різні аспекти їхнього комунікативного стилю, емоційного стану та когнітивних проявів у текстовій взаємодії.

Розширена множина ознак охоплює як традиційні метрики (тональність, емоційна валентність, частотність негативних/позитивних одиниць), так і нові інтегративні параметри, спрямовані на оцінку динамічних змін у текстовій структурі. До таких параметрів, зокрема, належать різка зміна емоційної тональності між реченнями, наявність погрозливих висловлювань, кількість граматичних або орфографічних помилок (які можуть сигналізувати про когнітивну дезорганізацію) та індикатори агресії за допомогою вбудованих або навчених моделей. Такі ознаки дають змогу створити більш об'ємне уявлення про автора повідомлення, що є критично важливим у контексті передбачення соціально небезпечної поведінки.

Ключовою особливістю методу є використання комбінованих індикаторів, що поєднують лексичні, емоційні та структурно-синтаксичні характеристики в єдиний профіль. Наприклад, поєднання високого рівня агресії із великою кількістю граматичних помилок і відсутністю позитивних емоцій у NRC може вказувати на високий ризик імпульсивної реакції або втрати самоконтролю. Аналогічно, виявлення різкого переходу між позитивною та негативною тональністю речень може свідчити про нестабільність афективного стану.

Особливу роль у межах даного підходу відіграє модуль аналізу смайлів та емотиконів, які, хоч і належать до неформальних засобів комунікації, однак виступають важливими маркерами емоційної експресії. Зменшення кількості таких символів, або їхня невідповідність загальному тону повідомлення, може служити додатковим індикатором емоційного виснаження або апатії. Водночас, надмірне використання негативних або агресивних емодзі також є потенційно тривожним сигналом.

Опис кожної ознаки систематизовано в таблиці 2.1, де вказано її функціональне призначення, тип значення та приклади реалізації. Така структура дозволяє не лише формалізувати кожен параметр у термінах

машинного навчання, але й забезпечити інтерпретованість результатів для фахівців у сфері психології та психіатрії.

Узагальнюючи вищенаведене, можна стверджувати, що побудова розширеної множини ознак створює передумови для багатовимірного аналізу користувацького контенту під час скринінгового оцінювання ризику соціально небезпечної поведінки осіб з ознаками ПТСР. Запропонований підхід також відкриває можливості для адаптивного налаштування систем моніторингу, що реагують не лише на явно виражені погрози, а й на приховані патерни емоційної деградації, втрати когнітивного контролю або прояви соціального відчуження.

Для формалізації розширення ознакового простору введемо вектор z_i , що для повідомлення m_i об'єднує модельну оцінку ПТСР, вектор оцінок супутніх розладів, емоційний контекст та додаткові лінгвістичні й поведінкові характеристики з таблиці 2.1. Емоційний канал у програмній реалізації формується моделлю roberta-base-go_emotions [131].

$$z_i = \psi(x_i, p_{PTSD,i}, p_{MD,i}, e_i, b_i) \in X'. \quad (2.15)$$

У дослідженій експериментальній конфігурації до індексу ризику (2.16)–(2.18) входять три складові вектора z_i : рішення щодо ознак ПТСР, рішення щодо інших психічних розладів та індикатор негативного емоційного контексту. Складову b_i (ознаки табл. 2.1) передбачено в розширеному профілі повідомлення як напрям подальшого розвитку методу; до розрахунку індексу в цій роботі вона не входить і експериментально не оцінювалась (п. 4.6). Сигнатура оператора методу: $T_3: (X' \times \mathbb{R}_+)^N \rightarrow [0, 1]^N$, де $N = N_u$ – кількість повідомлень у профілі користувача.

Для часової оцінки вводиться рухоме вікно $I_u(t, W) = \{i : t - W \leq t_{u,i} \leq t\}$. Базова складова ризику визначається як частка повідомлень з ознаками ПТСР у поточному вікні:

$$R_{base(u,t)} = \left(\frac{1}{|I_u(t,W)|} \right) \cdot \sum_{i \in I_u(t,W)} \mathbb{1}[p_{PTSD,i} \geq \tau_{PTSD}]. \quad (2.16)$$

Два додаткові коефіцієнти враховують коморбідність і негативний емоційний контекст. Щоб усунути неоднозначність початкового запису, модифікатори формалізовано через індикаторні функції: q_c – частка ПТСР-повідомлень, для яких виявлено щонайменше один супутній розлад; q_n – частка ПТСР-повідомлень з негативним емоційним контекстом.

$$J_c = \mathbb{1}[q_c > 0,30], \quad J_n = \mathbb{1}[q_n > 0,50], \quad (2.17)$$

$$R_{w,t} = \min\{1, R_{base(u,t)} + 0,2 \cdot J_c + 0,3 \cdot J_n\}. \quad (2.18)$$

Порогові значення 0,30 і 0,50 та коефіцієнти 0,2 і 0,3 є параметрами запропонованої експериментальної конфігурації, а не універсальними клінічними константами. Вибір значень обґрунтовано такими міркуваннями: сума приростів 0,5 не дає модифікаторам самим по собі вивести користувача з малою базовою часткою ПТСР-повідомлень на максимальний рівень ризику, тож визначальною лишається базова складова; приріст за негативний емоційний контекст (0,3) більший за приріст за коморбідність (0,2), оскільки емоційний контекст визначається однією моделлю, а коморбідність – п'ятьма бінарними моделями, помилки яких накопичуються; пороги 0,30 і 0,50 відповідають правилам «понад третину» і «понад половину» ПТСР-повідомлень у вікні. Чутливість індексу до цих параметрів і до ширини вікна W у роботі не досліджувалась (п. 4.6). Виходом методу є часовий ряд $R_{w,t}$ та розширений профіль ознак, який використовується для аналізу динаміки ризикових патернів.

Роль методу в досягненні мети: методи 1 і 2 забезпечують точність на рівні повідомлення, а метод розширення множини ознак – розширення можливостей оцінювання ризиків, тобто оцінку ризику на рівні користувача. Він переводить рішення методів рівня повідомлення та емоційний контекст у часовий індекс $R_{w,t} \in [0, 1]$ за (2.16)–(2.18); індекс адитивний, тому його значення пояснюване через складові (2.18), а вплив помилки окремого повідомлення обмежений усередненням у вікні (2.23).

2.5. Взаємозв'язок методів у комплексному аналізі

Оператори розроблених методів T_1 , T_2 і T_3 розв'язують різні за природою задачі: T_1 формує кількісну оцінку прояву ПТСР для окремого повідомлення, T_2 – вектор оцінок і рішень щодо K супутніх психічних розладів для того самого повідомлення, а T_3 – часовий ряд індексу ризику для користувача за сукупністю його повідомлень. Об'єднання цих методів у комплексний аналіз потребує окремого формального опису, оскільки визначаються не лише самі оператори, а й спосіб узгодження їх просторів входів і виходів, порядок виконання, умови активації поглиблених етапів та правила передавання проміжних результатів між методами. У цьому підрозділі наведено таку формалізацію, визначено два режими функціонування комплексу, досліджено вплив порогових параметрів і помилок окремих методів на інтегральний результат та встановлено властивості комплексу, які використовуються далі під час алгоритмічної реалізації (розділ 3) та експериментальної перевірки (розділ 4).

У комплексному аналізі виокремлено два рівні. На рівні повідомлення оператори діють на представленні $x = \varphi(m) \in X$, сформованому за (2.2): T_1 реалізує відображення $X \rightarrow [0,1]$ з подальшим пороговим рішенням (2.3)–(2.4), T_2 – відображення $X \rightarrow [0,1]^K$ із вектором рішень за індивідуальними порогами (2.14), а оператор E формує емоційний контекст повідомлення $e \in \mathcal{E}$. На рівні користувача оператор T_3 діє не на окремому представленні x , а на часово впорядкованій послідовності розширених векторів (2.15), побудованих для профілю P_u (2.1), і повертає часовий ряд $R_{u,t}$ за (2.16)–(2.18). Області визначення та області значень методів зведено в таблиці 2.2.

З таблиці 2.2 випливає, що область значень T_1 (відрізок $[0,1]$) не збігається з областю визначення T_2 (простір ознак X), а вхід T_3 формується не одним повідомленням, а множиною розширених векторів усього профілю. Тому формально комплексний аналіз не є композицією виду $T_3(T_2(T_1(x)))$.

Таблиця 2.2

Простори входів і виходів методів комплексного аналізу

Оператор	Рівень	Вхідні дані	Вихідні дані	Параметри	Споживач виходу
T_1 (метод 1)	повідомлення	$x = \varphi(m) \in X$	$p_{PTSD} \in [0,1]$ $\hat{y}_{PTSD} \in \{0,1\}$	θ_1, τ_{PTSD}	умова активації T_2
T_2 (метод 2)	повідомлення	$x = \varphi(m) \in X$	$p_{MD} = (p_1, \dots, p_K) \in [0,1]^K$ $\hat{y} \in \{0,1\}^K$	$\theta_2, \tau_1^*, \dots, \tau_K^*$	складова z_i (2.15); індикатор коморбідності (2.17)
E (емоційний контекст)	повідомлення	$m \in \mathcal{T}$	$e \in \mathcal{E}$; індикатор негативної тональності	θ_E	складова z_i (2.15); індикатор негативності (2.17)
ψ (розширення ознак)	повідомлення	$x, p_{PTSD}, p_{MD}, e, b$ (табл. 2.1)	$z \in X'$	—	вхід T_3
T_3 (метод 3)	користувач	$\{(z_i, t_i)\}, i = 1, \dots, N_u$	$\{R_{u,t}\} \subset [0,1]$	W , пороги 0,30 і 0,50, ваги 0,2 і 0,3	візуалізація трендів; висновок про ризик

Коректно спочатку сформувані паралельні модельні оцінки для повідомлення, а потім використати їх як частину розширеного профілю:

$$A(m) = (T_1(\varphi(m)), T_2(\varphi(m)), E(m)); \quad (2.19)$$

$$z = \psi(\varphi(m), A(m)); \quad T_3: \{z_i, t_i\} \rightarrow R_{u,t}.$$

Тут z – розширений вектор (2.15) в експериментальній конфігурації, тобто без складової b ; t_i – момент публікації повідомлення.

Подання (2.19) визначає повний (паралельний) режим, у якому всі оператори рівня повідомлення обчислюються для кожного повідомлення незалежно один від одного. Результат T_1 у цьому режимі є не входом T_2 , а

рівноправною складовою профілю $A(m)$: обидва методи аналізують те саме представлення $x = \varphi(m)$. Повний режим використовується для експериментального оцінювання кожного методу окремо (сценарії 1–3 у п. 3.7), оскільки дає змогу отримати оцінки T_2 для всіх повідомлень, а не лише для тих, у яких виявлено ознаки ПТСР.

Для практичного застосування, відповідно до загальної схеми на рис. 2.1, використовується умовно-послідовний режим, у якому поглиблений аналіз супутніх розладів активується лише після перевищення порогу ПТСР. Введемо індикатор активації

$$g(m) = \mathbb{1}[p_{PTSD} \geq \tau_{PTSD}] = \hat{y}_{PTSD}. \quad (2.20)$$

Тоді профіль повідомлення в умовно-послідовному режимі має вигляд:

$$A^c(m) = \left(T_1(\varphi(m)), g(m) \cdot T_2(\varphi(m)), E(m) \right), \quad (2.21)$$

де добуток $g(m) \cdot T_2(\varphi(m))$ означає, що за $g(m) = 0$ вектор оцінок супутніх розладів приймається нульовим, а самі моделі T_2 не запускаються. Режими (2.19) і (2.21) збігаються на підмножині повідомлень із $g(m) = 1$ і відрізняються лише для повідомлень без ознак ПТСР. Оскільки індикатор коморбідності у (2.17) обчислюється як частка саме ПТСР-повідомлень, для яких виявлено щонайменше один супутній розлад, обидва режими дають тотожний часовий ряд $R_{u,t}$; умовно-послідовний режим лише зменшує обчислювальні витрати. Це узгоджує вимогу практичної реалізації (аналіз громадян із виявленими ознаками ПТСР) із вимогою повноти експериментальної перевірки.

Жорсткий поріг у (2.20) може призводити до втрати повідомлень із граничними оцінками (наприклад, $p_{PTSD} = 0,49$ при $\tau_{PTSD} = 0,5$). Тому для практичної реалізації індикатор активації доцільно послабити зоною невизначеності: метод T_2 активується за умови $p_{PTSD} \geq \tau_{PTSD} - \delta$, де δ – ширина зони (наприклад, $\delta = 0,1$), а внесок таких повідомлень до індикатора коморбідності може зважуватися оцінкою p_{PTSD} . У повному режимі (2.19) ця проблема відсутня, оскільки всі оператори обчислюються для кожного

повідомлення; отже, можлива втрата граничних повідомлень є властивістю режиму економії обчислень, а не самого методу.

Порогові параметри методів є незалежними: поріг τ_{PTSD} належить процедурі прийняття рішення T_1 (2.4), пороги τ_k^* визначаються для кожної бінарної моделі T_2 окремо за індексом Юдена (2.14), а порогові значення 0,30 і 0,50 у (2.17) є параметрами часової агрегації. Важливою властивістю комплексу є те, що T_3 використовує результати методів рівня повідомлення лише через індикаторні функції, а не через неперервні оцінки. Позначивши для повідомлення m_i

$$\begin{aligned} s_i &= \mathbb{1}[p_{PTSD,i} \geq \tau_{PTSD}], \quad c_i = \mathbb{1}[\exists k: p_{k,i} \geq \tau_k^*], \\ n_i &= \mathbb{1}[e_i \in \mathcal{E}^-], \end{aligned} \quad (2.22)$$

де $\mathcal{E}^-$ – підмножина негативних емоційних станів, індекс ризику (2.16)–(2.18) можна подати як функцію лише бінарних величин: $R_{u,t} = F(\{s_i, c_i, n_i : i \in I_u(t, W)\})$. Наслідком є інваріантність T_3 до калібрування моделей: оцінки p_{PTSD} , p_k та e формуються моделями різних архітектур (DeBERTa, DistilBERT, RoBERTa) і не є порівнянними між собою як ймовірності, однак після пороговування вони входять до індексу ризику на однакових підставах. Заміна будь-якої з моделей або перевизначення її порогу не потребує зміни інших методів чи формул агрегації.

Разом з тим комплекс успадковує помилки методів рівня повідомлення. Нехай π – справжня частка повідомлень з ознаками ПТСР у вікні $I_u(t, W)$, а TPR_1 і FPR_1 – чутливість і частка хибнопозитивних рішень T_1 . Тоді математичне сподівання базової складової ризику (2.16) становить

$$\mathbb{E}[R_{base}(u, t)] = \pi \cdot TPR_1 + (1 - \pi) \cdot FPR_1, \quad (2.23)$$

тобто хибнопозитивні рішення T_1 завищують індекс ризику навіть для користувачів без ознак ПТСР, а хибнонегативні – занижують його та, в умовно-послідовному режимі, блокують аналіз супутніх розладів. Помилки T_2 впливають на індекс лише через індикатор коморбідності J_c і лише для ПТСР-повідомлень, а їх вплив обмежено ваговим коефіцієнтом 0,2. Усереднення в

межах вікна (2.16) послаблює вплив помилки окремого повідомлення пропорційно до $\frac{1}{|I_u(t,W)|}$, тому стійкість T_3 зростає зі збільшенням кількості повідомлень у вікні. Із (2.23) також впливає порядок експериментальної перевірки: якість T_1 і T_2 оцінюється спочатку і незалежно (сценарії 1 і 2), а T_3 перевіряється як механізм інтеграції та часової агрегації вже валідованих складових (сценарій 3), що відповідає підходу, прийнятому в розділі 4.

Узгодженість методів забезпечує також наскрізну пояснюваність результатів. На рівні повідомлення рішення T_1 пояснюється внеском окремих токенів і вагами уваги (п. 3.8, п. 4.2), рішення T_2 – вектором оцінок за окремими розладами з індивідуальними порогами. На рівні користувача індекс ризику (2.18) є адитивним: його значення декомпонується на базову частку ПТСР-повідомлень, внесок коморбідності $0,2 \cdot J_c$ та внесок негативного емоційного контексту $0,3 \cdot J_n$, тому для будь-якого моменту t можна вказати, які саме складові й які повідомлення сформували підвищений ризик.

Обчислювальні витрати комплексу визначаються кількістю нейромережових інференсів. Для користувача з N_u повідомленнями в умовно-послідовному режимі виконується N_u інференсів T_1 , N_u інференсів моделі емоційного контексту та $K \cdot \sum_{i=1}^{N_u} g(m_i)$ інференсів бінарних моделей T_2 , тобто

$$\Omega(u) = N_u \cdot (\omega_1 + \omega_E) + K \cdot \omega_2 \cdot \sum_{i=1}^{N_u} g(m_i), \quad (2.24)$$

де $\omega_1, \omega_2, \omega_E$ – вартість одного інференсу відповідної моделі (для трансформерних моделей вона пропорційна квадрату довжини послідовності токенів). Часова агрегація (2.16)–(2.18) для рухомого вікна виконується за один прохід упорядкованого профілю і має лінійну складність $O(N_u)$, тому не є обмежувальним чинником. У повному режимі другий доданок зростає до $K \cdot \omega_2 \cdot N_u$, що й обумовлює використання цього режиму лише для експериментів.

Таким чином, комплексний аналіз побудовано як узгоджену систему трьох методів із такими властивостями: 1) чітко визначені інтерфейси –

кожен метод споживає або представлення повідомлення $x = \varphi(m)$, або розширений профіль, а не «сирий» вихід іншого класифікатора; 2) умовна активація поглиблених етапів після перевищення порогу ПТСР без втрати тотожності підсумкового індексу ризику; 3) інваріантність часової оцінки до калібрування моделей завдяки порогованню на рівні повідомлення; 4) прогнозоване поширення помилок, що обґрунтовує послідовність експериментальної перевірки; 5) декомпозованість індексу ризику, яка забезпечує пояснюваність результату на рівні користувача. Ці властивості використано під час алгоритмічної реалізації методів (п. 3.1–3.4) та побудови сценаріїв експериментального дослідження (п. 3.7).

2.6. Висновки до розділу 2

У розділі сформовано теоретичні положення та математичну основу запропонованих методів інтелектуального аналізу користувацького текстового контенту для виявлення проявів ПТСР, супутніх психічних розладів та оцінювання ризику соціально небезпечної поведінки.

Удосконалено метод аналізу прояву ПТСР в користувацькому контенті. На відміну від проаналізованих підходів до автоматизованого виявлення ПТСР за текстовими даними [17–23], у запропонованому методі акцентовано контекстні залежності, специфічні для ПТСР, із урахуванням позиційної складової представлення тексту. Додатковою відмінністю є формування спеціалізованої навчальної вибірки, у якій нецільовий клас містить не лише тексти без ознак психічних розладів, а й тексти з проявами інших психічних розладів. Це забезпечує орієнтацію методу не лише на виявлення ознак ПТСР, а й на їх відмежування від семантично близьких проявів інших психічних станів. Метод формалізовано оператором (T_1), результатом якого є кількісна оцінка прояву ПТСР та відповідне порогове рішення.

Удосконалено метод виявлення психічних розладів за аналізом користувацького контенту. На відміну від багатокласового підходу, за якого

повідомленню примусово ставиться у відповідність один домінуючий клас [124], запропонований метод допускає незалежне формування позитивних рішень щодо декількох психічних розладів. Для цього задачу декомпозовано на множину задач бінарної класифікації за принципом бінарної релевантності [125, 126], для кожного розладу формується окрема навчальна вибірка, що містить у нецільовому класі також прояви інших розладів, та визначається індивідуальне порогове значення [129]. Виходом методу (T_2) є вектор кількісних оцінок та рішень щодо наявності кожного з досліджуваних психічних розладів без необхідності вибору лише одного домінуючого стану.

Вперше розроблено метод розширення множини ознак користувацького контенту для попередження соціально небезпечної поведінки громадян із посттравматичним стресовим розладом. На відміну від підходів, що ґрунтуються на аналізі окремих лінгвістичних, емоційних або поведінкових індикаторів [81–97], запропонований метод формує інтегрований цифровий профіль користувача шляхом поєднання оцінки прояву ПТСР, вектора оцінок супутніх психічних розладів, емоційного контексту повідомлень та часових характеристик цифрової активності. Поведінкову складову формалізовано через аналіз динаміки зазначених ознак у рухомому часовому вікні та побудову індексу ризику. Результатом (T_3) є часовий ряд індексу ризику, що дає змогу аналізувати не лише поточний стан окремого повідомлення, а й зміну ризикових патернів користувача в часі.

Основні наукові результати розділу опубліковано в [158–166, 169].

РОЗДІЛ 3. АЛГОРИТМІЧНА РЕАЛІЗАЦІЯ ТА МЕТОДИКА ЕКСПЕРИМЕНТАЛЬНОГО ДОСЛІДЖЕННЯ

3.1. Алгоритмічна реалізація методу аналізу прояву ПТСР у користувачькому контенті

Запропонований підхід до вирішення задачі виявлення проявів ПТСР в користувачькому текстовому контенті передбачає два етапи: препроцесинг і токенизацію та нейромережевий аналіз текстового контенту. Центальною ідеєю підходу є застосування контекстно-орієнтованої нейромережевої моделі, яка, опираючись на структуру трансформера, здатна враховувати глибокі семантичні та емоційні залежності в тексті.

Підхід перетворює вхідний текст за допомогою навченої контекстно-орієнтованої нейромережевої моделі з токенизатором у числову оцінку прояву ПТСР у користувачькому контенті.

На першому етапі здійснюється препроцесинг та токенизація текстового контенту. Контент перевіряється на довжину та на непустоту. Однак, розділові знаки, смайли та решта ознак тексту зберігаються, оскільки вони можуть нести додаткову потрібну інформацію, яка свідчить про наявність ПТСР. Токенизація здійснюється токенизатором, на якому була навчена нейронна мережа. Емодзі та пунктуація не видаляються на етапі препроцесингу і подаються токенизатору разом зі словами; токенизатор SentencePiece моделі DeBERTa-v3 відображає їх у підсловні одиниці свого словника, тому вони входять до контекстного представлення та враховуються механізмом уваги нарівні з лексичними токенами. Окремої обробки емодзі метод не передбачає; їх внесок у рішення може бути оцінений засобами пояснення результатів (п. 4.2).

На другому етапі здійснюється нейромережевий аналіз прояву ПТСР в користувачькому контенті, для виявлення якого буде використано контекстно-орієнтовану нейромережу архітектури трансформер.

Запропонований підхід ґрунтується на використанні контекстно-орієнтованої трансформерної моделі для аналізу текстів з метою виявлення ознак ПТСР. Обробка вхідного контенту здійснюється з урахуванням потенційно інформативних мовних ознак, включно з розділовими знаками та іншими стилістичними елементами, що зберігаються під час препроцесингу. Токенізація виконується із застосуванням засобів, сумісних зі структурою навченої моделі, що забезпечує коректне представлення тексту в її вхідному просторі. Функціонування моделі орієнтоване на інтерпретацію семантичних залежностей у тексті, що дозволяє отримати модельну оцінку належності повідомлення до класу з ознаками ПТСР. Підхід є формалізованим способом автоматизованого аналізу психоемоційного стану на основі мовленнєвих характеристик тексту.

Застосування методу до повідомлення складається з таких кроків: 1) перевірка, що текст непорожній і не перевищує допустимої довжини; 2) токенизація та формування представлення повідомлення за (2.2); 3) обчислення модельної оцінки прояву ПТСР оператором T_1 за (2.3); 4) порогове рішення за (2.4); 5) за потреби обчислення внесків токенів у рішення засобами пояснення (п. 4.2). Параметри оператора T_1 отримують формуванням набору даних і донавчанням моделі за алгоритмами 3.1–3.3 (п. 3.2).

3.2. Нейромережеві моделі для виявлення ПТСР

3.2.1. Нейромережева модель DeBERTa

Як контекстно-орієнтованої нейромережевої моделі буде використано модель DeBERTa [120] у варіанті DeBERTa-v3-small [140, 141] із шістьма шарами енкодера, архітектура якої наведена на рис. 3.1. Нейромережева архітектура DeBERTa є покращеною модифікацією моделі BERT, що реалізує декілька критично важливих удосконалень у механізмах представлення тексту, обробки контексту та моделювання залежностей. Її структура включає п'ять логічно взаємопов'язаних груп шарів, кожен з яких

виконує специфічну функціональну роль у процесі перетворення текстових даних на векторизоване представлення з подальшим формуванням модельної оцінки наявності ознак ПТСР.

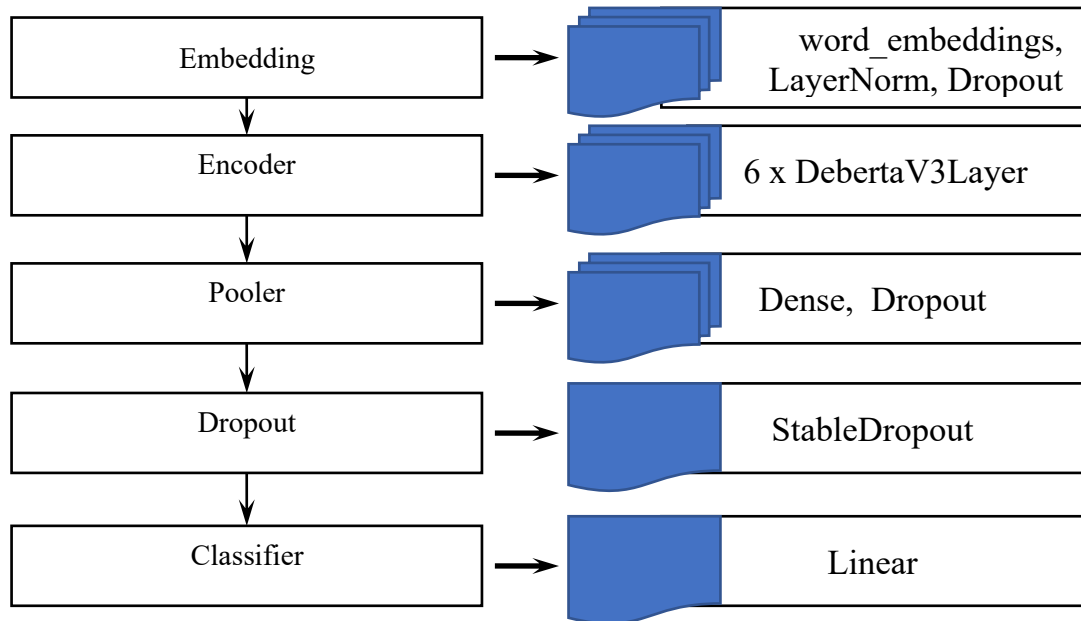


Рис. 3.1 – Архітектура моделі DeBERTa

Шар «Embedding» включає словникові ембединги та нормалізацію шару. На початковому етапі відбувається формування вхідних векторів за допомогою шару «Embedding», який містить лексичні ембединги слів та позицій. На відміну від традиційних моделей, DeBERTa реалізує роздільне моделювання семантичних і позиційних ембедингів. Такий підхід дозволяє моделі більш точно кодувати контекстуальні залежності, оскільки позиційна інформація обробляється окремо і не змішується з лексичними ознаками на ранньому етапі, що суттєво підвищує здатність до генералізації.

Множина шарів «Encoder» містить 6 шарів «DebertaV3Layer», кожен з них має компоненти «attention», «intermediate» та «output». Цей компонент моделі складається з багатьох шарів і реалізує основну обробку тексту. В ньому реалізовані механізми трансформера, які дозволяють моделі мати високу здатність до генералізації. Завдяки багаторазовій взаємодії між словами та контекстом, цей шар допомагає моделі навчитися складним залежностям і

контекстам, що дозволяє їй краще впоратися з новими або невідомими даними. Замість традиційного механізму уваги, DeBERTa використовує роздільну увагу, яка розглядає слова і позиції окремо. Виявлено, що окреме виконання компонентів «attention» для слова та позиції забезпечує збільшення обсягу сприйнятої взаємодії між словами та контекстом, що, в свою чергу, допомагає моделі аналізувати складні залежності і контексти, які притаманні задачі виявлення ознак ПТСР у користувацькому текстовому контенті [120]. Також цій моделі характерно використання відносного позиційного зміщення замість абсолютного, що дозволяє моделі краще враховувати контекстні залежності на різних відстанях завдяки логарифмічному групуванню відносних відстаней між токенами.

Шар «Pooler» агрегує отримані контекстуальні вектори та формує єдиний вектор представлення тексту, який подається на вхід класифікатора. Остаточну класифікацію здійснює лінійний шар «Classifier», який генерує модельну оцінку належності тексту до класу з проявами ПТСР. Шар «Dropout», вбудований у структуру моделі, виконує функцію регуляризації, зменшуючи ризик перенавчання моделі під час тренування на обмежених або дисбалансованих даних.

Отже, архітектурні особливості моделі DeBERTa, зокрема реалізація механізму роздільної уваги та використання відносного позиційного зміщення, дозволяють забезпечити обробку контекстно залежної інформації на рівні слів і позицій окремо. Це сприяє формуванню точніших представлень тексту, що є релевантними для задачі виявлення ознак ПТСР в користувацькому текстовому контенті. Структура моделі підтримує узгоджену трансформацію вхідного тексту у векторний простір з подальшою класифікацією за допомогою лінійного шару. Вбудовані механізми регуляризації знижують ймовірність перенавчання, що є важливим у контексті застосування моделі до даних із обмеженою кількістю позитивних прикладів.

3.2.2. Нейромережева модель з квантовим шаром

Як додатковий пошуковий експеримент, не пов'язаний з моделлю DeBERTa і не включений до складу методу аналізу прояву ПТСР, досліджено окрему повнозв'язну нейромережу з квантовим шаром. Архітектура нейромережі, представлена на рисунку 3.2, реалізує багатошаровий підхід до класифікації текстового контенту з метою виявлення ознак ПТСР. Мережа починається з вхідного шару, який приймає векторизований текст розмірності 2000. Цей вектор відповідає за попередньо оброблене уявлення тексту, яке може бути сформоване за допомогою TF-IDF, векторів частот або іншої методики кодування текстових ознак.

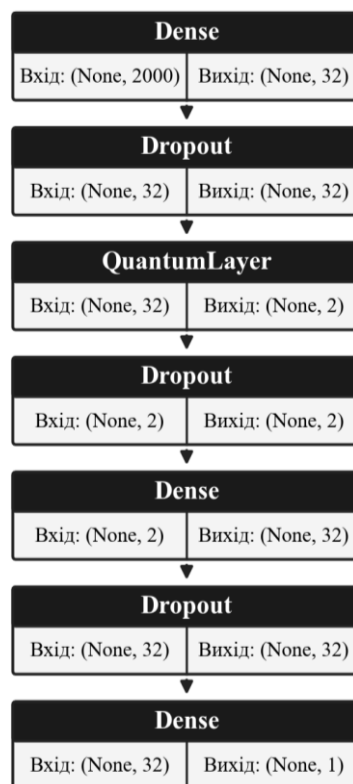


Рис. 3.2 – Архітектура нейромережі для виявлення проявів ПТСР у текстовому контенті

Перший щільний «Dense» шар виконує лінійне перетворення вхідного вектора до простору меншої розмірності, що дозволяє зменшити обчислювальну складність та виділити найбільш інформативні ознаки. Після

цього застосовується регуляризаційний механізм Dropout, який стохастично обнуляє частину активацій на цьому рівні, що сприяє підвищенню здатності моделі до узагальнення та зменшує ризик перенавчання. Ймовірність Dropout досліджено в діапазоні 0,6–0,75 (п. 4.2.1); такі значення вищі за типові 0,2–0,5 і зумовлені малим обсягом навчальної вибірки.

Центральним компонентом моделі є експериментальний квантовий шар, який здійснює нелінійне перетворення ознак на основі квантового обчислення. Квантовий шар має на меті моделювання складних взаємозв'язків у високорозмірних просторах за допомогою квантових станів та операцій. Шар симулюється засобами бібліотеки Cirq і перетворює вектор розмірності 32 на вектор розмірності 2 (рис. 3.2).

Після обробки квантовим шаром застосовується ще одна серія з щільного шару та Dropout, що дозволяє відновити та стабілізувати ознаки, отримані після квантового перетворення. Останній щільний шар перетворює результат у скаляр, який інтерпретується як модельна оцінка наявності ознак ПТСР у вхідному текстовому зразку. Це дозволяє реалізувати бінарну класифікацію. У дослідженій конфігурації Dropout застосовано і до виходу квантового шару розмірності 2; за ймовірності 0,7 обидва його елементи одночасно обнуляються приблизно в половині кроків навчання ($0,7 \times 0,7 = 0,49$), що є недоліком цієї конфігурації та однією з причин трактувати її результати як попередні.

Отже, модель із квантовим шаром розглядається в роботі як пошуковий експеримент: її результати наведено в п. 4.2.1, а до складу методу аналізу прояву ПТСР вона не входить.

Запропонована методика виявлення ПТСР у текстових даних наведена на рис. 3.3. Спершу необхідно підготувати робочі дані. Набір даних для виявлення ПТСР у текстових даних буде збірним із декількох наборів даних, та потребує балансування та попередньої обробки, що необхідно для ефективного навчання нейромережі. Схема формування набору даних детально описана в п. 3.5.1.

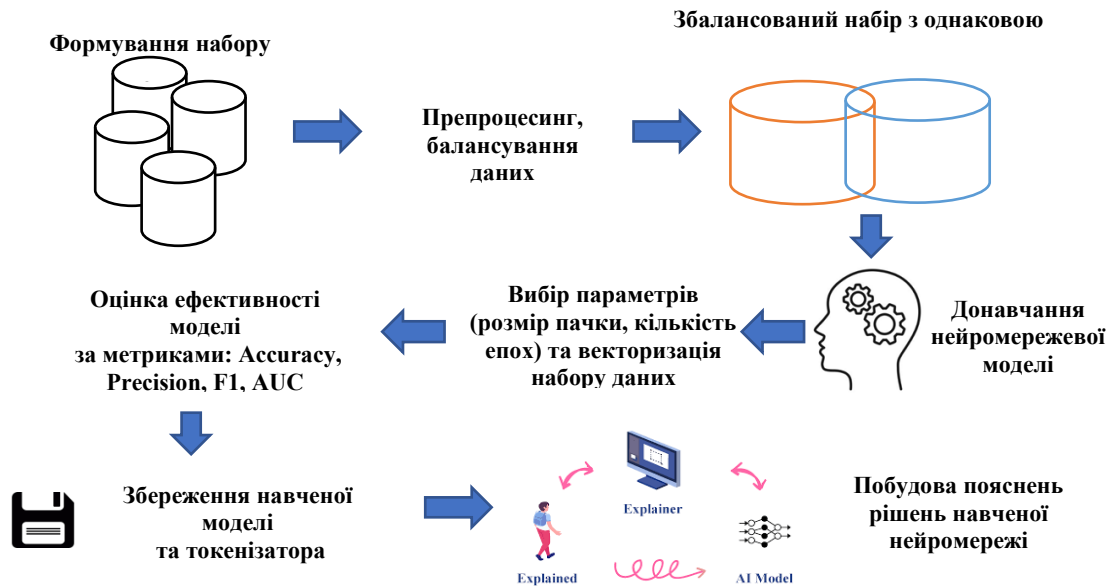


Рис. 3.3 – Кроки для нейромережевого виявлення ПТСР у текстових даних

Збалансований набір даних подається на вхід нейромережі для її подальшого донавчання. На цьому ж етапі здійснюється векторизація та поділ набору даних на навчальну та валідаційну вибірки у співвідношенні 80 % до 20 % відповідно. Також тут задаються параметри навчання, у вигляді розміру пачки та кількості епох.

Після завершення процесу донавчання нейромережі архітектури трансформер відбувається оцінка ефективності моделі за метриками: Accuracy, Precision, F1, AUC. Після обчислення метрик модель та токенизатор зберігаються.

Останнім кроком є пояснення рішень, отриманих за допомогою навченої нейромережі-трансформера. Для цього будуть використані інструменти пояснення моделі, що призначені для роботи з нейромережами архітектури трансформер.

Для створення якісного, чистого та збалансованого набору даних для навчання моделі виявлення ПТСР можна використати алгоритм 3.1:

Алгоритм 3.1. Формування та балансування набору даних

Vxid: множина вихідних текстових наборів даних $\{DS_1, \dots, DS_j\}$

- 1: Об'єднати всі набори даних в єдиний корпус.
- 2: Очистити тексти (вилучити зайві символи, привести до нижнього регістру).
- 3: Розмітити записи за класами:
 - ПТСП
 - інші розлади та тексти без розладів
- 4: Збалансувати класи:
 - якщо класи нерівні за обсягом, застосувати випадкове зменшення більшого класу до обсягу меншого

Vxid: збалансований набір даних DS_{bal}

На першому етапі всі наявні джерела текстових даних об'єднуються в єдиний корпус. Дані проходять попередню обробку: видаляються зайві символи, здійснюється нормалізація тексту (приведення до нижнього регістру, очищення від URL, спеціальних символів, цифр тощо). Кожному прикладу призначається мітка: ПТСП або інші психічні розлади. У разі дисбалансу класів застосовується одна з технік балансування, наприклад, зменшення більшості класу (*undersampling*). У результаті формується збалансований набір даних. Випадкове скорочення більшого класу застосовано свідомо: методи синтетичного передискретизування (SMOTE, ADASYN) оперують у просторі числових ознак і не породжують коректних природномовних текстів, а після об'єднання джерел співвідношення класів становить близько 2:1 і усувається без істотної втрати різноманітності даних (п. 3.5.1). Для методу виявлення психічних розладів дисбаланс і нечіткість меж класів додатково враховано функціями втрат *focal loss* і *dice loss* (п. 3.6).

Перетворення текстових даних у формат, придатний для подачі до трансформерної моделі, та розділення на навчальну та валідаційну вибірки здійснюється перетвореннями Алгоритму 3.2.

Алгоритм 3.2. Токенізація та поділ даних

Вхід: збалансований набір даних DS_{bal}

- 1: Завантажити токенизатор, на якому навчено трансформерну модель.
- 2: Виконати токенизацію текстів (відображення f).
- 3: Поділити набір:
 - 80 % → навчальна вибірка
 - 20 % → валідаційна вибірка
- 4: Зберегти вибірки для подальшого навчання.

Вихід: навчальна та валідаційна вибірки у векторизованому вигляді

На цьому етапі застосовується токенизатор моделі DeBERTa-v3, яке перетворює текст на послідовності токенів. Текст представлений у вигляді числових векторів з урахуванням уваги та маскування. Отримані дані діляться у співвідношенні 80:20 – більша частина використовується для навчання, менша – для перевірки якості моделі.

Донавчання попередньо навченого трансформера на спеціалізованому наборі для виявлення ПТСР проводиться за алгоритмом 3.3.

Алгоритм 3.3. Донавчання моделі

Вхід: токенизовані навчальна та валідаційна вибірки

- 1: Завантажити попередньо навчену трансформерну модель.
- 2: Задати параметри навчання:
 - розмір пачки
 - кількість епох
 - швидкість навчання.
- 3: Навчити модель на навчальній вибірці, мінімізуючи функцію втрат (3.1).
- 4: Після кожної епохи обчислити метрики на валідаційній вибірці.
- 5: Після завершення навчання:
 - зберегти параметри моделі θ_1
 - зберегти токенизатор

Вихід: донавчена модель T_1

Попередньо навчена трансформерна модель DeBERTa-v3-small завантажується з відкритого репозиторію моделей [140]. Задаються параметри навчання, такі як розмір пачки, кількість епох та швидкість навчання. Модель навчається на навчальній вибірці, а після кожної епохи оцінюється на валідаційній. Цей процес дозволяє адаптувати загальну мовну модель до специфіки ПТСР-контекстів у текстах. Після завершення навчання параметри моделі та токенизатор зберігаються для подальшого застосування.

Запропонований підхід до виявлення ПТСР у текстових даних базується на поетапній підготовці, балансуванні та обробці набору даних, що забезпечує умови для ефективного навчання нейромережі. Використання трансформерної архітектури дозволяє враховувати складні контекстні залежності завдяки впровадженню відносного позиційного кодування та роздільного аналізу уваги до слів і їх позицій. Логарифмічне групування відносних відстаней між токенами сприяє більш точному відображенню взаємозв'язків на різних відстанях у тексті. Включення до навчального набору зразків із проявами інших психічних розладів дозволяє покращити здатність моделі розрізняти симптоми ПТСР від інших схожих проявів. Загалом, описана методика формує системний процес, який забезпечує якісне навчання і перевірку моделі з подальшим поясненням прийнятих рішень.

3.3. Алгоритмічна реалізація методу виявлення психічних розладів за аналізом користувацького контенту

Наведена у п. 2.3 формальна модель реалізується у вигляді послідовності процедур формування бінарних наборів, тонкого налаштування окремих моделей та отримання вектора рішень за індивідуальними порогами. На схемах і в алгоритмах нижче збережено деталі програмної реалізації, тоді як математичні відображення повторно не вводяться.

Як базову модель DLM для бінарних класифікаторів множини M обрано DistilBERT [132]. Такий вибір обумовлений тим, що стандартні моделі BERT схильні до надмірного підлаштування під окремі слова або маркери, що в текстах із психоемоційним навантаженням може призводити до упереджених висновків, тоді як DistilBERT завдяки меншій глибині та навчанню через дистиліацію має стабільнішу узагальнювальну здатність на нових даних від різних авторів з різними стилями комунікації, а також потребує менших обчислювальних ресурсів для навчання п'яти окремих моделей.

На рис. 3.4 зображено схему послідовності обробки даних для отримання множин Df та M .

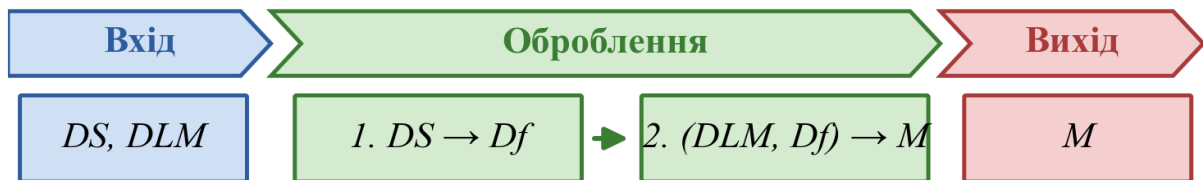


Рис. 3.4 – Схема обробки даних для отримання множин Df та M

На кроці 1 (див. рис. 3.4), за допомогою *Алгоритму 3.4*, отримується множина наборів даних Df (відображення f_2), необхідних для тонкого налаштування множини моделей глибокого навчання M .

Алгоритм 3.4. Реалізація відображення f_2 (формування бінарних наборів даних)

Вхід: DS

1: Для $i = 1$ до K

2: Ініціалізація:

- $\text{target_class} \leftarrow$ підмножина DS із записами, мітка яких (id_{DSO}) відповідає цільовому класу
- $\text{non_target_classes} \leftarrow$ підмножина DS із записами інших класів

- $\text{non_target_samples} \leftarrow \emptyset$
 - $\text{sample_size_per_source} \leftarrow 0$
 - $\text{target_count} \leftarrow |\text{target_class}|$
 - 3: Визначити обсяг вибірки з кожного нецільового класу:

$$\text{sample_per_source} = \text{target_count} / K$$
 - 4: Сформувати вибірку з нецільових класів (non):
 - $\forall id_{\text{DSO}} \in \text{non}$
 - обрати випадкову підвибірку обсягом $\text{sample_size_per_source}$ і присвоїти їй мітку «нецільовий»
 - додати підвибірку до $\text{non_target_samples}$.
 - 5: Об'єднати отримані вибірки:

$$Df_i = \text{target_class} \cup \text{non}$$
 - 6: Перемішати записи у наборі даних Df_i
 - 7: Зберегти сформований набір даних Df_i
 - 8: **Кінець циклу**
- Вихід:** набори даних Df_i
-

На кроці 2 (див. рис. 3.4), за допомогою Алгоритму 3.5, тонко налаштовується множина моделей глибокого навчання M (відображення f_3).

Алгоритм 3.5. Донавчання множини моделей M

- Вхід:** набори даних Df_i , базова модель DLM , токенізатор f_1
- 1: **Для** $i = 1$ **до** K
 - 2: Ініціалізація: завантажити попередньо навчену модель DLM з токенізатором f_1 та набір даних Df_i
 - 3: Випадково поділити Df_i на навчальну (Df_train) і тестову (Df_test) вибірки
 - 4: Виконати попереднє оброблення Df_train і Df_test токенізатором f_1
 - 5: Токенізувати тексти та сформувати мітки у Df_train і Df_test

- 6: Донавчити модель μ_i для психічного розладу id_{DsO_i}
- 7: Оцінити модель за метриками на Df_train і Df_test
- 8: Визначити поріг для психічного розладу:
 - отримати передбачення моделі μ_i на Df_train
 - обчислити оцінки $p = \sigma(\ell)$ сигмоїдною функцією активації, де ℓ – логіт моделі
 - визначити оптимальний поріг за індексом Юдена:
- 9: Оцінити модель із порогом τ за метриками на Df_train і Df_test
- 10: Зберегти донавчену модель μ_i та поріг τ
- 11: **Кінець циклу**

Вихід: донавчені моделі M , оптимальний поріг τ для кожної μ_i

Приклад створення набору даних для навчання Df_1 з цільовою групою для психічного розладу d_1 наведено на рис. 3.5.

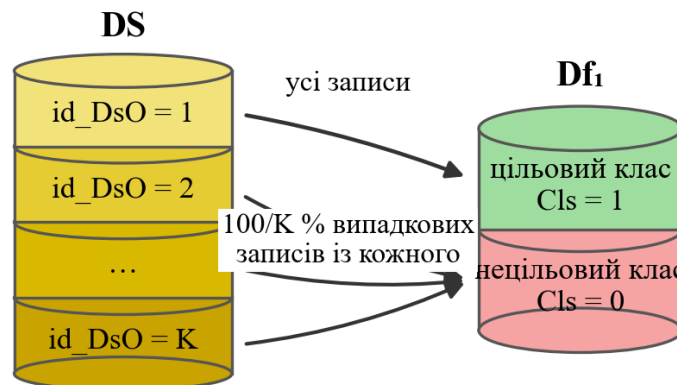


Рис. 3.5 – Приклад формування набору даних Df_1 для розладу d_1 ($id_{DsO} = 1$)

На рис. 3.6 зображено схему обробки даних для отримання рішення про наявність чи відсутність психічних розладів.

На кроці 1 (рис. 3.6), за допомогою відображення f_1 з вхідного тексту t отримуємо токенований текст tkn .



Рис. 3.6 – Схема обробки даних для отримання рішення про наявність чи відсутність певного психічного розладу

На кроці 2 (рис. 3.6), за допомогою Алгоритму 3.6, отримується результат застосування M до токенозованого тексту tkn : (відображення f_4) – рішення про наявність чи відсутність психічних розладів r .

Алгоритм 3.6. Отримання вектора оцінок r і вектора рішень $\hat{y}$ (відображення f_4)

Вхід: моделі M , токенозований текст tkn

- 1: Для $i = 1$ до K
- 2: Отримати оцінку p психічного розладу d_i моделлю μ_i для tkn
- 3: Якщо $p \geq \tau$ то $\hat{y}_i \leftarrow 1$, інакше $\hat{y}_i \leftarrow 0$
- 4: $r_i \leftarrow p$
- 5: Кінець циклу

Вихід: $r, \hat{y} = (\hat{y}_1, \dots, \hat{y}_K)$

Наведено метод автоматизованого виявлення психічних розладів на основі текстового контенту користувачів, зокрема з соціальних мереж. Замість традиційного підходу з багатокласовою класифікацією, запропоновано багатоміткову класифікацію (MLC), яка допускає виявлення декількох розладів одночасно, не втрачаючи другорядні, але потенційно критичні ознаки. Модель будується шляхом декомпозиції MLC на набір бінарних задач, де для кожного типу розладу створюється окрема модель.

3.4. Алгоритмічна реалізація методу розширення множини ознак користувацького контенту

Запропонований метод базується на використанні нейромережових моделей для багаторівневого аналізу текстів, що дає змогу не лише ідентифікувати потенційно травматичний контент, а й досліджувати його часову динаміку у взаємозв'язку з проявами емоційної дисрегуляції та супутніх психічних розладів. У результаті формується скринінговий індекс ризику, який може слугувати основою для раннього виявлення груп підвищеного ризику та подальшого розроблення превентивних стратегій підтримки психічного здоров'я. Схема методу наведена на рис. 3.7.

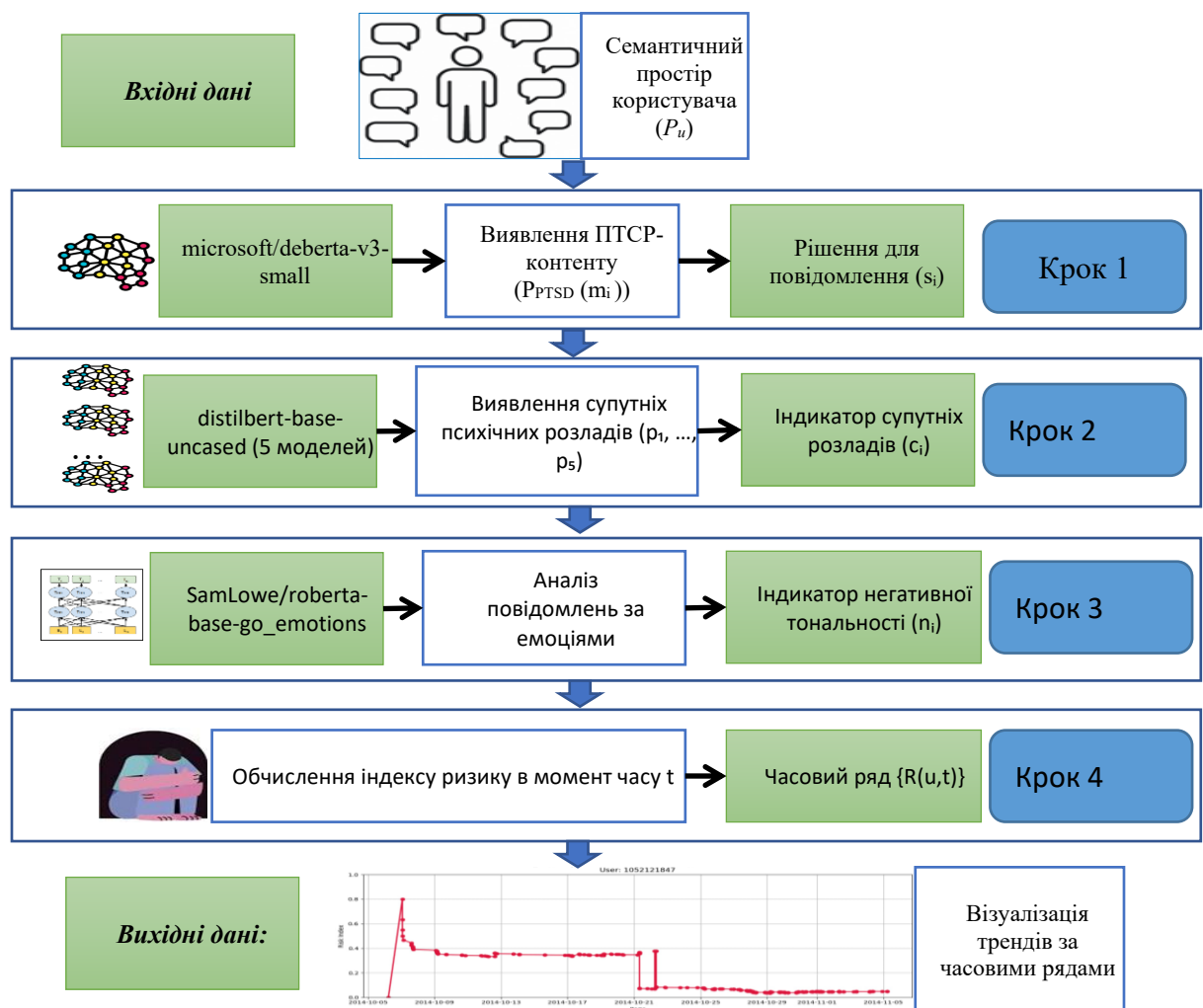


Рис. 3.7 – Схема методу розширення множини ознак користувацького контенту для попередження соціально небезпечної поведінки

Метод починається з використання семантичного простору користувача як вхідних даних, що відображає його активність у соціальних мережах та містить усі повідомлення для аналізу з датою та часом, тобто профіль користувача P_u , формалізований у (2.1).

На кроці 1 здійснюється аналіз повідомлень на наявність ознак ПТСР, що дозволяє ідентифікувати потенційно травматичний контент та виділити його як основу для подальшої оцінки. Для кожного повідомлення m_i нейромережева модель обчислює модельну оцінку прояву ознак ПТСР відповідно до (2.3).

Якщо $p_{\text{PTSD}}(m_i) \geq \tau_{\text{PTSD}}$ (в експерименті $\tau_{\text{PTSD}} = 0,5$), повідомлення вважається таким, що містить ознаки ПТСР ($s_i = 1$), інакше $s_i = 0$. Для виявлення ознак ПТСР у повідомленні використано нейромережу BERT-подібної архітектури, що є результатом попередніх досліджень.

На кроці 2 відбувається аналіз повідомлень на наявність ознак супутніх психічних розладів, що забезпечує більш комплексну оцінку психічного стану користувача і дозволяє врахувати коморбідні прояви. У дослідженні аналізуються такі види психічних розладів: anger, anxiety, depression, narcissistic, panic. Відповідно, для кожного повідомлення формується вектор оцінок супутніх розладів розмірності $K = 5$ – за кількістю досліджуваних розладів.

Для визначення кожного з психічних розладів використано нейромережеві моделі BERT-подібних архітектур, що є результатами попередніх досліджень. Кроки 1 і 2 не є складовими методу розширення множини ознак: на них лише використовуються виходи методів 1 і 2 (оцінка p_{PTSD} та вектор p_{MD}). Власний зміст методу становлять кроки 3 і 4 – формування емоційного контексту та часова агрегація розширеного вектора ознак за (2.15)–(2.18).

На кроці 3 проводиться аналіз повідомлень за емоційним контекстом, що дає змогу оцінити характер і тональність емоційних проявів у текстах. В межах

дослідження важливим є виявлення саме негативної тональності. Індикатор негативної тональності входить до емоційного контексту e_i у (2.15).

Визначення тональності виконується з використанням нейромережі roberta-base-go_emotions [131].

Крок 4 виконує часовий аналіз текстового профілю, що дозволяє досліджувати динаміку змін ризикової поведінки за обрахованим індексом ризику та емоційних реакцій користувача у межах визначених часових вікон. Для цього кожному повідомленню прив'язується мітка часу t_i , що дозволяє виконати ковзний аналіз у межах рухомого вікна розміром W днів (в експерименті $W = 30$ днів). Для кожного користувача u обчислюється базовий рівень ризику як відношення кількості повідомлень, що містять ПТСР-контент, до загальної кількості повідомлень у межах вікна відповідно до (2.16).

Додатково враховуються два модифікуючі фактори: Boost-коморбідності з ваговим коефіцієнтом 0,2 активується, якщо понад 30 % ПТСР-повідомлень супроводжуються іншими психічними розладами, а Boost-негативності з ваговим коефіцієнтом 0,3 додається, якщо понад 50 % ПТСР-повідомлень у межах вікна мають негативний емоційний контекст. Індекс ризику для користувача u у момент часу t обчислюється за (2.17) – (2.18).

Індекс $R(u, t) \in [0, 1]$ – нормалізований скринінговий індекс ризику. Отримані часові ряди $\{R(u, t)\}$ дозволяють відстежувати індивідуальні тренди ризикових патернів користувачів, що забезпечує змогу раннього виявлення груп підвищеного ризику.

Алгоритм працює з часово впорядкованим профілем користувача P_u , формалізованим у (2.1). Для кожного повідомлення послідовно обчислюються оцінка ПТСР за (2.3), вектор супутніх розладів за моделями п. 2.3 та емоційний контекст моделлю roberta-base-go_emotions [131]. Отримані значення разом із додатковими ознаками утворюють розширений вектор (2.15).

Після цього для рухомого вікна W обчислюється базова частка ПТСР-повідомлень (2.16), індикатори коморбідності та негативності (2.17) і індекс ризику (2.18). У такий спосіб алгоритмічна реалізація безпосередньо

відповідає формалізації розділу 2 й не потребує повторного введення тих самих формул. Послідовність обчислень наведено в алгоритмі 3.7.

Алгоритм 3.7. Обчислення часового індексу ризику

Вхід: профіль користувача P_u (2.1), оператори T_1 , T_2 і модель емоційного контексту, пороги рішень методів 1 і 2 за (2.4) і (2.14), ширина вікна W

- 1: Для кожного повідомлення m_i профілю P_u обчислити модельну оцінку ПТСР (2.3), вектор оцінок психічних розладів (2.12) та емоційний контекст
- 2: Для кожного повідомлення m_i сформувати індикатори s_i , c_i , n_i за (2.22)
- 3: Для кожного моменту t публікації повідомлення сформувати вікно $I_u(t, W)$
- 4: Обчислити базову складову ризику за (2.16)
- 5: Обчислити індикатори коморбідності та негативного емоційного контексту за (2.17)
- 6: Обчислити індекс ризику $R_{u,t}$ за (2.18)

Вихід: часовий ряд індексу ризику $\{R_{u,t}\}$

В результаті обробки формуються вихідні дані у вигляді візуалізації трендів за часовими рядами, що інтегрує інформацію про ПТСР-контент, супутні психічні розлади та емоційний контекст, забезпечуючи комплексну оцінку потенційних ризиків для кожного користувача.

Запропонований метод забезпечує комплексний підхід до автоматизованої оцінки посттравматичної ризикової поведінки на основі аналізу текстів користувачів у соціальних мережах. Інтеграція кількох нейромережових моделей дозволяє одночасно виявляти ПТСР-контент, супутні психічні розлади та емоційний контекст повідомлень, що розширює можливості оцінювання. Завдяки побудові індивідуальних часових рядів індексу ризику

метод дає змогу досліджувати динаміку змін поведінкових патернів користувачів та своєчасно ідентифікувати групи підвищеного ризику. Таким чином, метод формує основу для подальшого створення систем підтримки психічного здоров'я та розроблення превентивних стратегій.

3.5. Набори даних та їх підготовка

3.5.1. Дані для виявлення ПТСР

Оскільки немає відкритих наборів даних з достатньою кількістю даних для навчання нейромережі, було сформовано збірний набір даних на основі існуючих наборів даних «Human Stress Prediction» та «Aya PTSD».

Так, набір даних «Human Stress Prediction» [122] містить дані, опубліковані в субредітах, пов'язані з психічним здоров'ям. Цей набір даних містить різні проблеми психічного здоров'я, якими люди розповідають про своє життя. Набір даних містить таку розмітку: «subreddit», «post_id», «sentence_range», «text», «label», «confidence», «social_timestamp» представляє заголовки для файлу Stress.csv. З цього набору даних будуть взяті дані, які у колонці «subreddit» мають значення «PTSD», що свідчать про наявність ПТСР. Така мітка є похідною від тематики сабредіту, а не клінічним підтвердженням діагнозу, тому в межах роботи йдеться про виявлення ознак ПТСР у тексті, а не про його діагностування.

У протилежну категорію будуть віднесені записи, що не містять ознак ПТСР (всі решта записи). Кількість цільових записів із «PTSD» у даному наборі даних становить 584. Кількість записів, що віднесені до протилежної категорії: 2254. За тематикою сабредитів записи протилежної категорії розподілено так (у відсотках від усіх записів набору, рис. 3.8): relationships – 19,5 %, anxiety – 17,7 %, domesticviolence – 11,1 %, assistance – 10,2 %, survivorsofabuse – 8,6 %, homeless – 5,9 %, almosthomeless – 2,8 %, stress – 2,3 %, food_pantry – 1,3 %. Отже, нецільовий клас утворюють дописи тематично суміжних спільнот, лише частина яких (anxiety, stress) стосується психічних розладів.

Оскільки в решті даних будуть записи не тільки здорових людей, а і записи з різноманітного роду психічних розладів, вважається, що краще будуть шукатись контекстні залежності, що відповідають саме за ПТСР, та буде менше плутанини із рештами психічних розладів. Однак, це може призвести до гірших показників точності при навчанні. Отриманий набір даних вийшов незбалансованим, цільовий клас майже в 4 рази менше за протилежний йому. Тому ці дані є потреба доповнити іншим набором даних. Схема формування набору даних наведена на рис. 3.8.

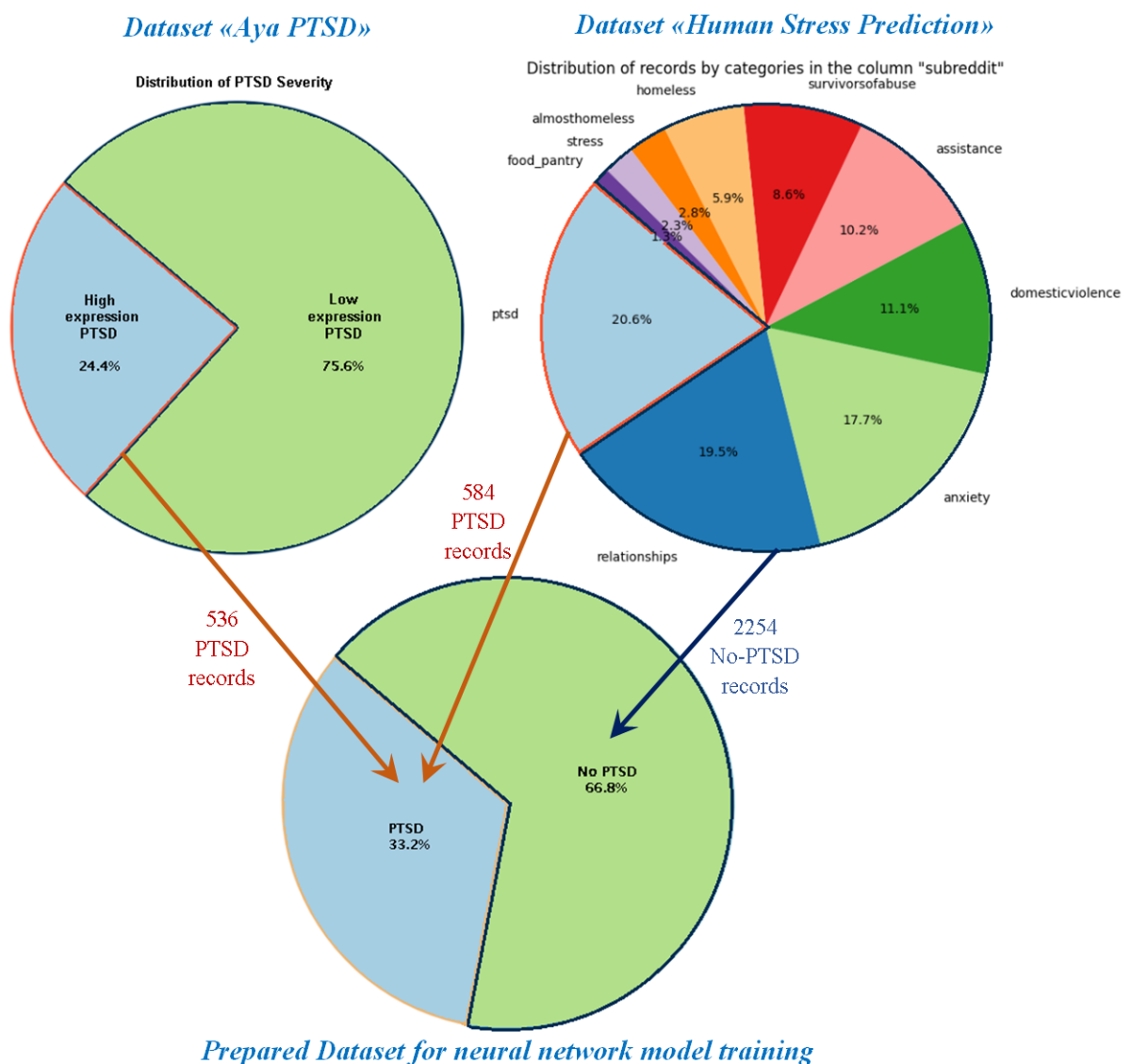


Рис. 3.8 – Розподіл записів у вихідних наборах даних та структура сформованого набору даних

Його буде доповнено набором даних «Ауа PTSD» [123], доступний на Kaggle, містить дані, які стосуються саме ПТСР. Він включає різні атрибути, такі як клінічні оцінки, соціально-демографічні характеристики тощо («ID», «PTSD Severity», «label», «text», «file_path», «response»), що робить його всеосяжним ресурсом для вивчення ПТСР. З цього набору даних будуть взяті записи, що говорять про рівень ПТСР атрибуту «PTSD Severity» більше 50 %. Поріг 50 % є параметром формування вибірки. Після доповнення даними, цільовий клас став містити 1120 записів (рис. 3.9).

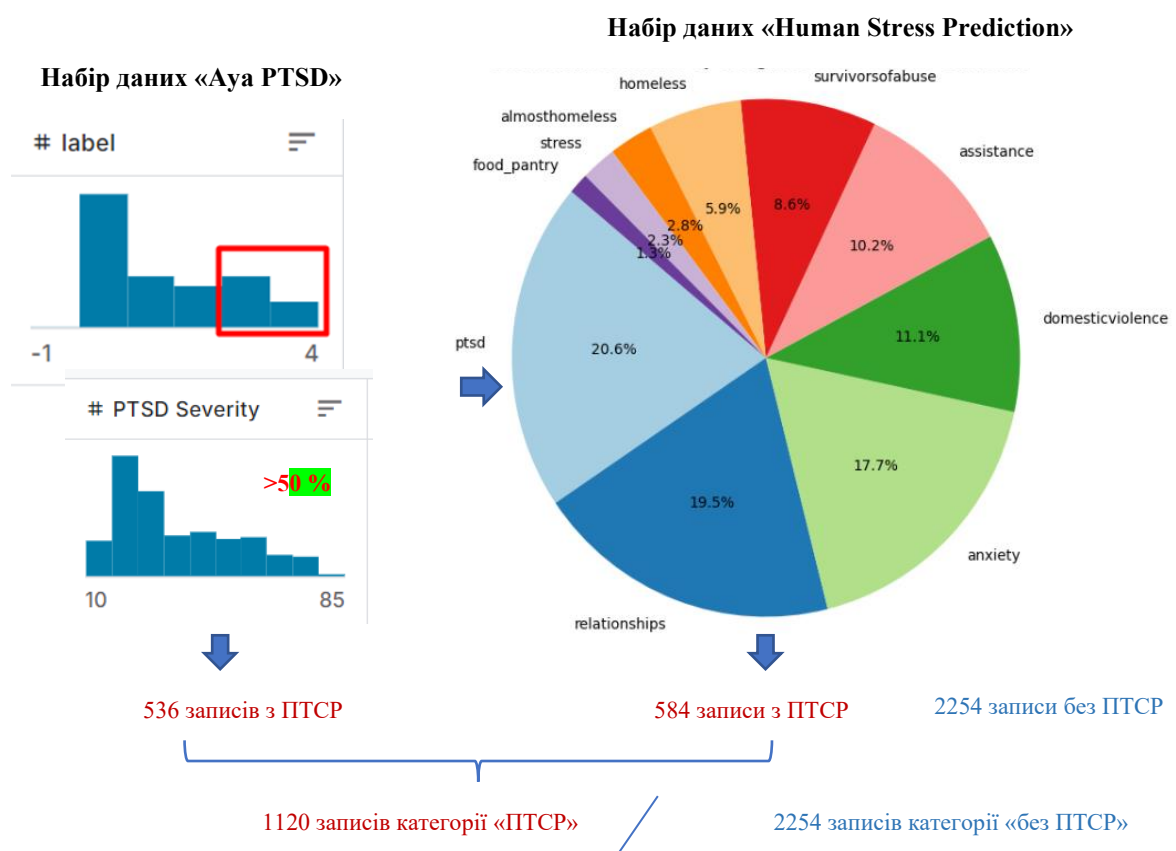


Рис. 3.9 – Схема формування збірного набору даних для навчання нейромережевої моделі

Таким чином було сформовано набір даних із 3374 записів, що складається із 2-х категорій: «ПТСР» та «Без ПТСР». Кількість записів категорії «ПТСР» становить 1120, а кількість записів протилежної категорії – 2254.

При цьому, під час навчання нейромережі буде використовуватись 2200 записів, а саме 1100 записів з категорії «ПТСР» та 1100 записів з категорії «Без ПТСР».

Для навчання нейромережі було сформовано збірний набір даних на основі двох існуючих джерел – «Human Stress Prediction» та «Aya PTSD». Перший містить записи з різними психічними станами, серед яких були виділені дані, що позначені як пов'язані з ПТСР, проте через незбалансованість класів виникла необхідність доповнення іншими даними. Другий набір даних, більш сфокусований на ПТСР, забезпечив додаткові записи, які дозволили збільшити кількість прикладів цільового класу. В результаті сформований набір даних містить дві категорії – з проявами ПТСР та без них – з достатньою кількістю збалансованих записів для навчання. Під час навчання моделі буде застосовано вибірку, яка забезпечує рівну кількість прикладів обох класів, що має сприяти покращенню якості класифікації.

3.5.2. Дані для виявлення психічних розладів

Для реалізації та тестування запропонованого підходу, виходячи з існуючих наборів даних «COMSYS-T1» [133] та «Depression: Reddit Dataset (Cleaned)» [134], далі розглядались наступні психічні розлади: d_1 – Anger/Intermittent Explosive Disorder, d_2 – Anxiety Disorder, d_3 – Depression, d_4 – Narcissistic Disorder, d_5 – Panic Disorder.

Набір даних «COMSYS-T1», складається із текстів із твітів, які містять потенційні психічні розлади особистості. Загальна кількість твітів – 785, з яких до розладу d_3 належить 208 зразків, розладу d_4 – 158 зразків, розладу d_1 – 154 зразки, розладу d_2 – 153 зразки та , розладу d_5 – 112 зразків.

Набір даних «Depression: Reddit Dataset (Cleaned)» містить очищений набір даних Reddit про розлад d_3 . Загальна кількість зразків у наборі даних – 7650. З них 3900 таких, що не містять ознак розладу d_3 та інших психічних розладів. Набір даних має клас «CC0: Суспільне надбання», тобто особа, яка

створила набір даних, передала його суспільному надбанню, відмовившись від усіх своїх прав на набір даних у всьому світі, включаючи всі пов'язані та суміжні права.

Для запобігання плутанини психічних розладів між собою та враховуючи, що в одному тексті можуть бути і прояви інших розладів з множини D , у нецільовій категорії записи формуються за певними правилами

- кількість дописів у нецільовій категорії відповідає, або наближається до цільової (похибка не більше 10 дописів);
- нецільова категорія складається із рівних пропорцій решти текстів із іншими видами психічних розладів та текстів що не містять таких проявів.

Схема побудови набору даних наведена на рис. 3.10

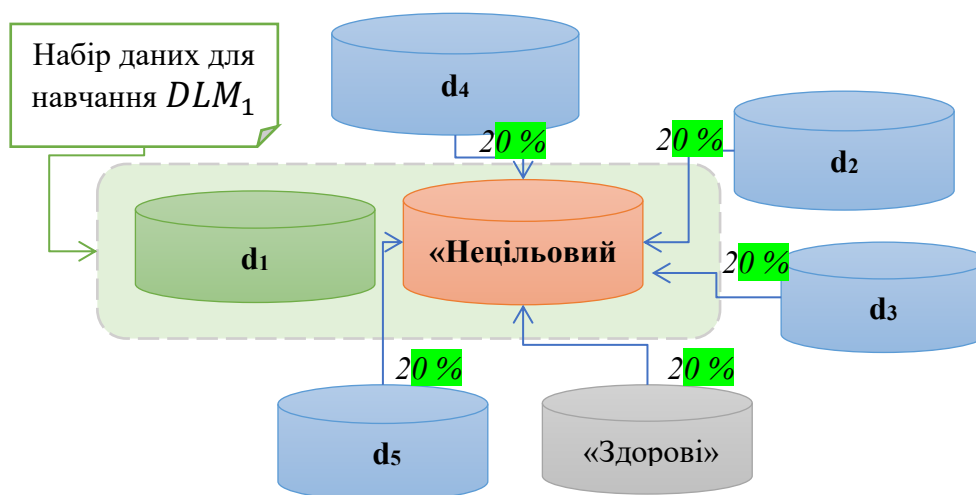


Рис. 3.10 – Схема побудови набору даних

Такий розподіл дозволяє виробити в моделях покращену здатність розрізняти специфічні ознаки для кожного розладу. Це зменшує ймовірність плутанини між розладами, оскільки модель вчиться розрізняти їхні характерні ознаки.

У рамках реалізації запропонованого підходу побудовано збалансовані набори даних на основі корпусів COMSYS-T1 та Depression: Reddit Dataset (Cleaned), що охоплюють тексти, пов'язані з проявами окремих психічних

розладів. Формування вибірок здійснювалося з урахуванням пропорційного представлення як цільових, так і нецільових категорій, з дотриманням принципів рівноваги та контролю за наявністю змішаних ознак. Така організація даних забезпечує структурну чіткість у відображенні психічних розладів і створює умови для навчання моделей, здатних до диференційованого розпізнавання їхніх ознак. Відповідне структурування вибірок сприяє зменшенню впливу потенційної надмірної кореляції між класами та покращує здатність моделей виявляти специфічні патерни кожного розладу.

3.5.3. Дані для часової оцінки ризикової поведінки

Запропонований у розділі 2 метод апробовано на наборі даних «Depression: Twitter Dataset + Feature Extraction», що опублікований на Kaggle [135].

Набір даних містить 20 000 позначених англійських твітів, зібраних для аналізу психічного здоров'я. Він включає дані про депресивні та недепресивні висловлювання користувачів соціальних мереж, призначені для досліджень у галузі обробки природної мови та виявлення психічних розладів. Також набір даних містить поле «post_created» і поле «user_id», що робить його придатним до часового аналізу семантичного простору користувача (User Semantic Space), тобто профілю P_u (2.1).

3.6. Метрики оцінювання та функції втрат

Метрики оцінювання точності MLC-класифікаторів. Очевидно, що для звичайних класифікаторів, які передбачають мітку єдиного класу, та MLC-класифікаторів, які формують на виході вектор міток, метрики відрізнятимуться. Зазвичай, для оцінки точності MLC класифікаторів застосовують (крім загальноприйнятих метрик для багатокласової

класифікації) такі метрики: Hamming loss (метрика базується на відстані Хеммінгу та визначається як відносна частота неправильних класифікацій) та 0/1-loss (метрика є часткою навчальних прикладів, при класифікації яких було допущено помилку хоча б для однієї мітки). Зауважимо, що метрика Hamming loss застосовується для методу BR, а метрика 0/1-loss може застосовуватися для методів CC та LP. Однак, метрика Hamming Loss не є придатною для використання в рамках нашого дослідження, оскільки наявні набори даних розмічено відповідно до задачі SLC. Водночас, структура анотацій не передбачає відсутність інших потенційних діагнозів у кожного пацієнта. Замість цього гарантується лише наявність конкретного розладу, який асоціюється з відповідною міткою. Таким чином, застосування метрик, орієнтованих на повноту міток, таких як Hamming Loss, є методологічно необґрунтованим, оскільки множинність можливих (але не відображених) міток не враховується у процесі розмітки.

Застосування різних функцій втрат – focal loss, dice loss та log loss – у межах експерименту обґрунтоване необхідністю забезпечити оптимальне навчання моделей в умовах класового дисбалансу, нечіткості меж класів та різної чутливості до помилок класифікації. Focal loss є модифікацією log loss, яка пригальмовує вплив легко класифікованих прикладів і фокусує навчання на складних зразках [136]. Це особливо важливо в задачах, де класи мають різну виразність або представлені нерівномірно, як у даному дослідженні.

Dice loss походить з сегментаційних задач, але є ефективною у класифікації при обмеженій кількості прикладів цільового класу [137]. Вона безпосередньо оптимізує F1-міру, що робить її релевантною, коли важливе не просто передбачення, а баланс між Precision і Recall. Log loss (крос-ентропія) залишається базовим стандартом у задачах бінарної класифікації, забезпечуючи стабільну конвергенцію та єдиний орієнтир для порівняння ефективності альтернативних функцій втрат [138].

Формально для бінарного класифікатора з оцінкою $p \in [0, 1]$ та істинною міткою $y \in \{0, 1\}$ використані функції втрат мають вигляд:

$$L_{\log} = -[y \cdot \log p + (1 - y) \cdot \log(1 - p)] \quad (3.1)$$

$$L_{\text{focal}} = -\alpha \cdot (1 - p_y)^\gamma \cdot \log p_y \quad (3.2)$$

$$L_{\text{dice}} = 1 - \frac{2 \cdot \sum_i p_i y_i + \varepsilon}{\sum_i p_i + \sum_i y_i + \varepsilon} \quad (3.3)$$

де $p_y = p$ при $y = 1$ та $p_y = 1 - p$ при $y = 0$; α і γ – параметри фокусування focal loss; ε – згладжувальна константа; підсумовування у (3.3) виконується за прикладами навчальної пачки.

Таким чином, включення всіх трьох функцій дозволяє оцінити адаптивність моделей до різних сценаріїв навчання й відібрати ті підходи, які найкраще відповідають особливостям кожної моделі з множини M .

Окрім наведеного, слушним є оцінювання бінарних класифікаторів, які лежать у основі методу BR, за метриками, притаманними для бінарної класифікації: *Accuracy* – частка правильно класифікованих прикладів від загальної кількості; *Precision* – частка правильних позитивних передбачень серед усіх передбачених позитивних випадків; *Recall* – повнота, частка правильних позитивних передбачень серед усіх реальних позитивних випадків та *F1-score* – середнє гармонійне між *Precision* та *Recall* [139]. Також, для доведення спроможності класифікатора, видобувати приховані залежності з навчального набору даних, потрібно навести графіки ROC-кривих та площі під ними (AUC).

Наведені вище показники обраховуються за параметрами матриці помилок: TP (True Positive) – кількість правильно передбачених позитивних випадків; TN (True Negative) – кількість правильно передбачених негативних випадків; FP (False Positive) – кількість хибно передбачених позитивних випадків; FN (False Negative) – кількість хибно передбачених негативних випадків.

Відповідно метрики визначаються як:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}. \quad (3.4)$$

$$\text{Precision} = \frac{TP}{TP+FP}, \quad \text{Recall} = \frac{TP}{TP+FN}. \quad (3.5)$$

$$F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \quad (3.6)$$

У контексті задач багатоміткової класифікації застосування класичних метрик точності потребує адаптації з огляду на специфіку векторної структури вихідних міток. Використання таких метрик, як Hamming loss та 0/1-loss, дозволяє охарактеризувати загальну якість класифікації, враховуючи відповідність кожної з міток. Водночас, у випадку, коли структура даних не гарантує повноти міткової анотації, застосування метрик, що базуються на припущенні про повну множину релевантних класів, є методологічно необґрунтованим. Саме тому в дослідженні акцент зроблено на функціях втрат, здатних адаптуватися до умов класового дисбалансу та нечітких кордонів між класами. Зокрема, використання focal loss, dice loss і log loss дозволяє забезпечити ефективне навчання моделей із урахуванням варіативності складності прикладів та різної вагомості класифікаційних помилок. Крім того, метрики, притаманні бінарній класифікації, використовуються для оцінювання складових класифікаторів у методі бінарної релевантності, що дає змогу здійснити більш детальне й обґрунтоване порівняння моделей. Побудова ROC-кривих та обчислення AUC забезпечують додаткову верифікацію здатності моделей виявляти релевантні ознаки в умовах багатовимірного простору ознак.

Для першого методу використовуються Accuracy, Precision, Recall/F1 та ROC-AUC; для другого – ті самі метрики для кожного бінарного класифікатора з подальшим macro-усередненням. Hamming Loss не використовується як основна метрика через неповноту багатоміткової анотації наявних вихідних наборів; це обмеження сформульовано вище. Для порівняння режимів навчання другого методу додатково досліджуються log loss, focal loss і dice loss [136-138].

3.7. Дизайн експериментальних досліджень

Експериментальну перевірку організовано у трьох сценаріях, що однозначно відповідають трьом методам і не змішуються з їх теоретичним описом.

Сценарій 1 – виявлення ПТСР. Порівнюються реалізації на основі DeBERTa та моделі з квантовим шаром; оцінювання виконується на валідаційній вибірці (20 % збалансованого набору) за Accuracy, Precision, F1 та ROC-AUC. Для DeBERTa додатково аналізується пояснюваність рішення засобами Transformers Interpret і BertViz [140-144].

Сценарій 2 – виявлення психічних розладів. Для кожної бінарної моделі застосовується наведена нижче процедура повторних розбиттів, а запропонований підхід порівнюється з багатокласовою та звичайною багатомітковою класифікацією [43, 145-147].

Пропонується така методика для оцінювання спроможності моделей M виявляти приховані залежності та виявляти психічні розлади. Для кожного набору даних Df_i та моделі μ_i :

Крок 1. Розбиття набору даних Df_i на навчальний та тестовий.

Крок 2. Тонке налаштування моделі μ_i .

Крок 3. Обчислення функції втрат тонкого налаштування.

Крок 4. Обчислення матриці помилок, метрик, ROC-кривої та AUC для навчальної вибірки.

Крок 5. Обчислення матриці помилок, метрик, ROC-кривої та AUC для тестової вибірки.

Крок 6. Повторення кроків 1–5 для 4 випадкових розбиттів (фолдів).

Крок 7. Розрахунок середніх значень метрик та максимальних відхилень від них.

Виходячи з чутливості предметної області задачі (здоров'я людей) та використаної багатоміткової класифікації пропонується використати такі метрики: *Accuracy*, *Precision*, *Recall* та *F1-score*.

Наступні експерименти направлені на порівняння запропонованого підходу з іншими. Пропонується порівняти запропонований підхід з іншими: багатокласовою класифікацією [43], звичайною багатомітковою класифікацією [145-147].

Для експерименту з багатокласовою класифікацією було використано набір даних «COMSYS-T1» у базовому вигляді, який є розміченим під багатокласову класифікацію. Для адаптації набору даних «COMSYS-T1» до багатоміткової класифікації було змінено спосіб представлення цільових міток. Оскільки початковий формат передбачав багатокласову класифікацію, де кожен зразок мав лише одну категорію, необхідно було перетворити мітки у формат, що дозволяє позначати кілька класів одночасно.

Це було реалізовано шляхом використання бінарного кодування кожної можливої мітки у вигляді вектора, де 1 означає наявність відповідного класу, а 0 – його відсутність. Оскільки множини класів не перетиналися, отримані вектори мали ортогональну структуру, що еквівалентно початковому багатокласовому представленню, але дозволяє працювати у багатомітковому форматі.

Було використано та навчено нейромережеву модель DistilBERT як багатокласового класифікатора для класів з множини D . Навчену модель DistilBERT для багатоміткової класифікації вищеперерахованих розладів, а також навчено множину M для кожного з 5 видів досліджуваних психічних розладів з множини D . Обґрунтування вибору DistilBERT наведено в п. 3.3.

Сценарій 3 – розширення множини ознак і часовий ризик. Для користувачів із доступними часовими мітками повідомлень будується часовий ряд $R_{w,t}$ за (2.18), після чого аналізуються піки й тренди разом із ПТСР-оцінками, супутніми станами та емоційним контекстом. Експеримент виконується на наборі, описаному у п. 3.5.3 [135].

3.8. Етичні аспекти, пояснюваність та межі автоматизованих рішень

Етичні аспекти та відповідальність за рішення в контексті аналізу психічних розладів на основі текстового контенту набувають особливої ваги через чутливість оброблюваної інформації та потенційний вплив рішень, що приймаються на її основі. Така діяльність пов'язана з безпосереднім втручанням у сферу приватного життя особи, що вимагає дотримання високих стандартів конфіденційності, етичності й професійної обережності [148, 149].

Аналіз психічних розладів на основі текстових даних вимагає врахування етичних аспектів, пов'язаних із конфіденційністю та відповідальністю за прийняті рішення. Застосування автоматизованих систем до даних конкретних осіб потребує правової підстави для обробки таких даних, а результати аналізу не можуть бути єдиною підставою для діагностичних висновків без залучення фахівців [150-152]. Важливо забезпечити достовірність і обережність інтерпретації результатів, обмежити передачу інформації третім сторонам відповідно до законодавства, а також уникати відтворення упереджень у моделях [153-155]. Технології слід використовувати як допоміжний засіб у комплексному процесі оцінки, що базується на професійному аналізі та дотриманні етичних норм [150, 151, 156].

У дослідженні використано лише відкриті знеособлені набори даних, опубліковані для наукового використання («Human Stress Prediction», «Aya PTSD», «COMSYS-T1», «Depression: Reddit Dataset (Cleaned)» з ліцензією CC0, «Depression: Twitter Dataset + Feature Extraction»). Виконано вторинний аналіз уже опублікованих наборів даних, тому окрема згода авторів дописів не отримувалась; ідентифікатори користувачів і дослівні тексти дописів у роботі не наводяться. Ідентифікатори користувачів застосовуються виключно як псевдонімні ключі для групування повідомлень у часові ряди; персональні дані не збираються і не зберігаються, результати роботи методів не передаються третім сторонам і не використовуються для постановки діагнозу [152, 153].

Результати моделей у роботі трактуються як автоматизовані індикатори для аналізу користувацького контенту, а не як самодостатній клінічний діагноз. Пояснюваність першого методу експериментально досліджується через візуалізацію внеску токенів та механізмів уваги.

Розроблені методи призначені для скринінгу мовних ознак у текстах і не є засобом клінічної діагностики чи прогнозування поведінки конкретної особи. Мітки використаних наборів даних відображають тематику онлайн-спільнот або розмітку їх упорядників, а не клінічно підтверджені діагнози; модельні оцінки не калібровано як ймовірності. Хибнопозитивне рішення може призвести до стигматизації людини, хибнонегативне – до пропуску потреби в допомозі, тому будь-яке практичне застосування потребує участі кваліфікованого фахівця в ухваленні рішення, правової підстави для обробки даних і контролю наслідків помилкових рішень.

Згідно з етичними принципами, процеси прийняття рішень алгоритмами, що використовують моделі глибокого навчання, повинні бути прозорими. Ця прозорість є важливою для осіб, на яких впливають ці рішення, оскільки вона дозволяє їм зрозуміти обґрунтування рішень, зроблених від їхнього імені, особливо коли це може стосуватися їхніх законних прав [149, 157].

Отже, вкрай важливо включити адекватні засоби та процеси в підходи на основі штучного інтелекту для уточнення процесів прийняття рішень у цих системах.

Крім того, методика, запропонована в цьому дослідженні для виявлення психічних розладів шляхом аналізу текстового контенту соціальних мереж, буде вдосконалена в майбутніх дослідженнях шляхом включення надійних процесів, які пояснюють користувачам рішення, прийняті моделями глибокого навчання.

3.9. Висновки до розділу 3

У розділі подано алгоритмічну реалізацію трьох методів без повторного введення їх математичних моделей; систематизовано набори

даних, правила підготовки вибірок, метрики, функції втрат і три незалежні сценарії експериментальної перевірки. Така структура забезпечує перехід від формалізації розділу 2 до експериментальних результатів розділу 4.

Алгоритмічно реалізовано запропонований у розділі 2 підхід до виявлення психічних розладів через аналіз користувацького контенту на основі багатоміткової класифікації з трансформацією даних, зокрема через метод бінарної релевантності, який має суттєві переваги. Він дозволяє не тільки ефективно виявляти найбільш виражені психічні розлади, але й уникати втрати менш яскраво виражених симптомів, що може мати вирішальне значення для своєчасного надання психологічної допомоги. Завдяки використанню MLC моделюється можливість одночасного присвоєння кількох міток, що є особливо важливим, коли симптоми можуть вказувати на кілька розладів. Крім того, врахування специфіки психічних розладів через коректне формування навчальних даних дозволяє підвищити точність ідентифікації та уникнути ситуацій, коли через взаємну неповноту міток втрачаються важливі ознаки. Проте залишаються деякі виклики, зокрема проблеми з коректним формуванням навчальних вибірок і відсутністю достатньої кількості відмічених даних, що негативно впливає на точність класифікації.

Значна увага приділяється формуванню набору даних, його балансуванню, попередній обробці тексту та коректній векторизації з урахуванням емоційних маркерів, таких як смайли та розділові знаки. Результати роботи нейромережі оцінюються за допомогою традиційних метрик якості класифікації. Завершальним етапом аналізу є пояснення прийнятих рішень моделлю, що підвищує її прозорість і довіру до висновків системи.

Визначено етичні умови та межі застосування методів (п. 3.8): результати трактуються як скринінгові індикатори, а не клінічний діагноз, і потребують участі фахівця в ухваленні рішень.

Алгоритмічну реалізацію запропонованих методів, підходи до формування наборів даних і програмні засоби опубліковано в працях [158–163, 167–169].

РОЗДІЛ 4. ЕКСПЕРИМЕНТАЛЬНЕ ДОСЛІДЖЕННЯ ТА АНАЛІЗ РЕЗУЛЬТАТІВ

4.1. Програмне середовище експериментів

З метою дослідження ефективності розробленого методу аналізу прояву ПТСР в текстовому контенті було створено відповідне програмне забезпечення у вигляді Notebook у хмарному середовищі Google Colab.

Дослідження ефективності було проведено за допомогою використання потужностей хмарного сервісу «Google Colab», який надає підтримку для різних бібліотек квантових обчислень, що дозволяє виконувати симуляції квантових алгоритмів та нейронних мереж. Для цього в дослідженні використовувалася бібліотека «Cirq».

Для експериментів другого методу програмну реалізацію виконано у вигляді IPython Notebook; п'ять бінарних моделей навчалися у середовищі CPU, а багатокласовий і багатомітковий класифікатори – з використанням TPU v2-8. Конкретні параметри навчання наведено безпосередньо у відповідних підрозділах результатів.

Зведення конфігурацій експериментів. Сценарій 1: модель microsoft/deberta-v3-small, донавчання протягом 3 епох на збалансованому наборі з 2200 записів (1760 навчальних і 440 валідаційних), 293 кроки на епоху (рис. 4.2), загальний час навчання близько 10 годин у середовищі Google Colab; окрема модель із квантовим шаром – до 100 епох із ранньою зупинкою, розмір пачки 64, розмір словника 2000, симуляція засобами Cirq. Сценарій 2: п'ять бінарних моделей distilbert-base-uncased, швидкість навчання $2e-5$; у варіанті без адаптивного порогу – розмір пачки 16 і 5 епох, у варіанті з адаптивним порогом – розмір пачки 4 і 4 епохи; поділ кожного набору на навчальну і тестову частини 80:20, чотири випадкові розбиття. Сценарій 3: моделі сценаріїв 1 і 2 та готова модель SamLowe/roberta-base-go_emotions, ширина вікна $W = 30$ днів. Набори даних і правила їх формування описано в п. 3.5, програмний код – у Додатку В.

4.2. Дослідження методу аналізу прояву ПТСР у користувацькому контенті

4.2.1. Дослідження нейромережевої моделі з квантовим шаром

Нейронна мережа з квантовим шаром, натренована на запропонованому наборі даних, показала такі результати за метриками: Accuracy 85 %, Precision 87 %, AUC 0,88. У таблиці 4.1 наведено порівняння результатів цих метрик для різних параметрів навчання на тому ж наборі даних з використанням квантового шару і без нього.

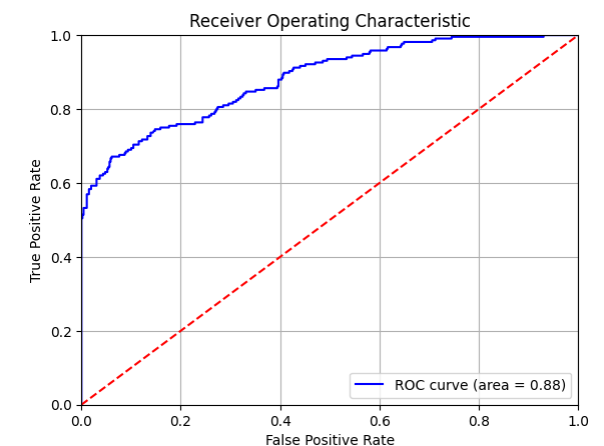
Таблиця 4.1

Показники метрик при різних налаштуваннях навчання на запропонованому наборі даних з квантовим шаром і без нього

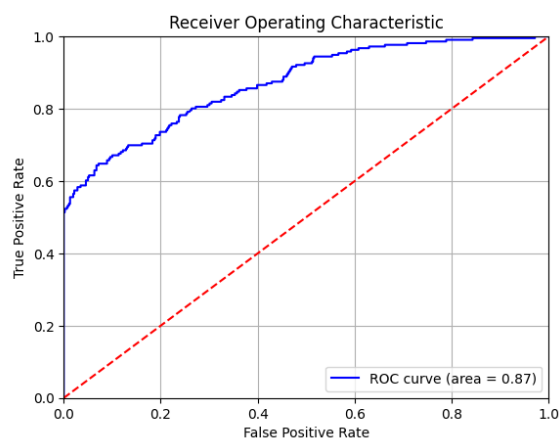
Параметри моделі	Accuracy (%)	Precision (%)	AUC
З квантовим шаром, частка Dropout 0,60	83	77	0,87
З квантовим шаром, частка Dropout 0,70	85	85	0,88
З квантовим шаром, частка Dropout 0,75	85	87	0,88
Без квантового шару, частка Dropout 0,60	84	86	0,87
Без квантового шару, частка Dropout 0,70	75	65	0,76
Без квантового шару, частка Dropout 0,75	82	79	0,82

Як показано в таблиці 4.1 та на рис. 4.1 (а)–(г), найкраща конфігурація з квантовим шаром (Dropout 0,75) має Accuracy 85 %, Precision 87 %, AUC

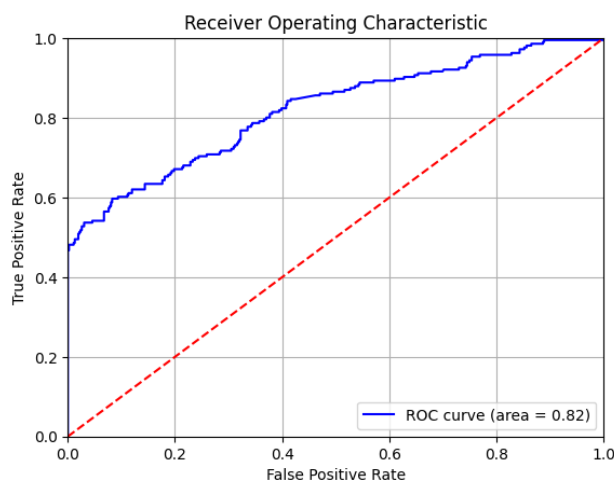
0,88, а найкраща конфігурація без квантового шару (Dropout 0,60) – Accuracy 84 %, Precision 86 %, AUC 0,87. За однакового значення Dropout 0,60 модель з квантовим шаром поступається моделі без нього (Accuracy 83 % проти 84 %, Precision 77 % проти 86 %), тому різницю між найкращими конфігураціями не можна однозначно приписати квантовому шару.



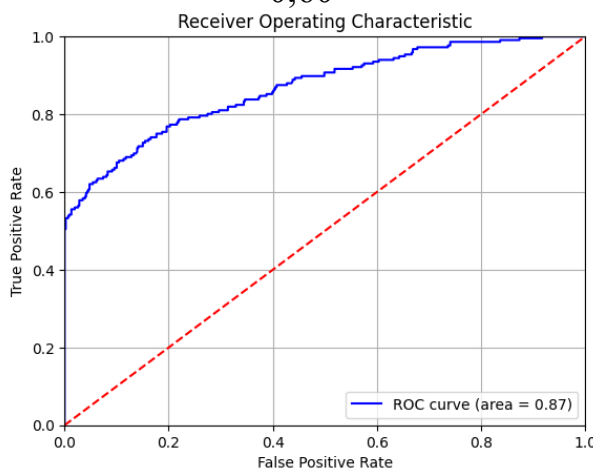
а) ROC-крива нейромережі з квантовим шаром, частка Dropout 0,75



б) ROC-крива нейромережі з квантовим шаром, частка Dropout 0,60



в) ROC-крива нейромережі без квантового шару, частка Dropout 0,75



г) ROC-крива нейромережі без квантового шару, частка Dropout 0,60

Рис. 4.1 – ROC-криві при різних параметрах нейромережових архітектур

Отже, для архітектури з квантовим шаром спостерігалось збільшення Accuracy і Precision на 1 відсотковий пункт та AUC на 0,01 порівняно з найкращою моделлю без квантового шару; статистична значущість цієї різниці не перевірялася. Різницю отримано за один запуск для конфігурацій з

різними значеннями Dropout, тому приписати її саме квантовому шару не можна.

Навчання нейромереж з параметрами таблиці 4.1 здійснювалось протягом 100 епох із застосуванням ранньої зупинки для захисту від перенавчання. Зберігалися лише найкращі моделі протягом усього процесу навчання, а не лише по завершенню окремих епох. Розмір навчальної пачки (batch) обрано 64, при розмірі словника 2000 токенів. Частку Dropout досліджено в діапазоні 0,6–0,75, вищому за типові значення 0,2–0,5, через малий обсяг навчальної вибірки (2200 записів) при розмірності вхідного вектора 2000, що потребує сильної регуляризації для запобігання перенавчанню.

Розроблений метод має обмеження, оскільки працює лише з англійськими текстами, але в разі наявності навчальних даних на інших мовах це обмеження може бути усунене. Майбутні дослідження будуть зосереджені на розширенні навчальної вибірки та вивченні альтернативних архітектур нейронних мереж для покращення точності виявлення ПТСР.

Отже, модель із квантовим шаром є пошуковим експериментом: її метрики (Accuracy 0,85) нижчі, ніж у моделі DeBERTa (Accuracy 0,934, п. 4.2.2), тому до складу методу аналізу прояву ПТСР вона не входить, а підтвердження ефекту квантового шару потребує повторних запусків з однаковими гіперпараметрами.

4.2.2. Дослідження нейромережевої моделі DeBERTa

Як переднавченої нейронної мережі архітектури трансформер була взята модель «microsoft/deberta-v3-small» з ресурсу «HuggingFace» [140]. Модель «DeBERTaV3» покращує оригінальну модель «DeBERTa» шляхом заміни моделювання мови масок (MLM) заміненим виявленням маркерів (RTD), більш ефективним для вибірки завданням попереднього навчання [141].

Цю модель донавчено на створеному та описаному вище наборі даних протягом 3-х епох. Статистику навчання за епохами наведено на рис. 4.2.

```
Epoch 1/3
100%|██████████| 293/293 [3:19:54<00:00, 40.94s/it]
Training loss: 0.4747
Epoch 2/3
100%|██████████| 293/293 [3:20:02<00:00, 40.97s/it]
Training loss: 0.3364
Epoch 3/3
100%|██████████| 293/293 [3:19:09<00:00, 40.78s/it]
Training loss: 0.2519
```

Рис. 4.2 – Статистика навчання

Загалом, навчена модель була протестована на валідаційній вибірці, яка показала такі значення за метриками: Accuracy 0,934, Precision 0,948, F1 0,933, AUC 0,872. ROC-криву для оцінки якості навчання нейромережі наведено на рис. 4.3.

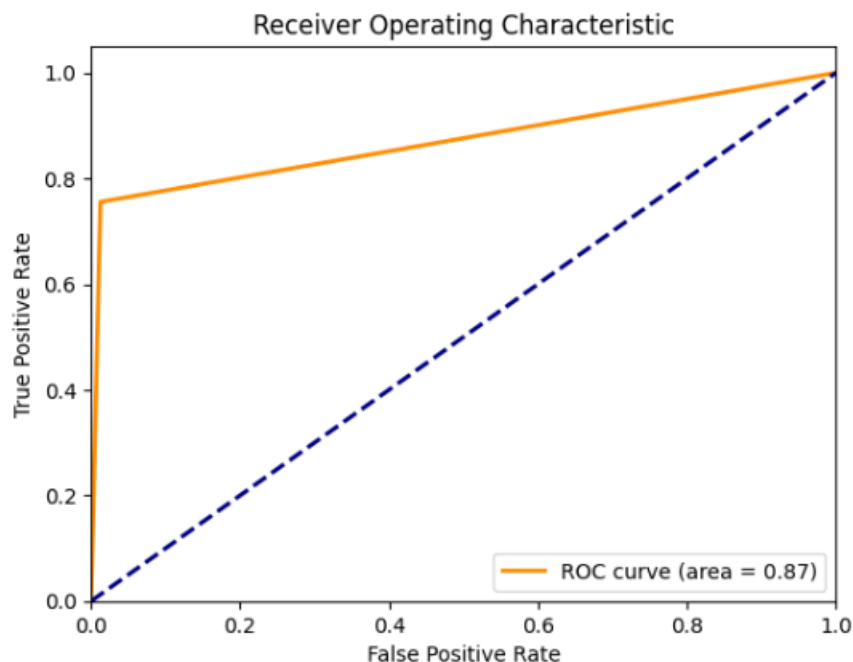


Рис. 4.3 – ROC-крива для оцінки якості навчання нейромережі

Значення Precision 0,948 означає, що з усіх зразків, які модель визначила як позитивні (ПТСР), 94,8 % дійсно були позитивними. Це

важливо у контексті ПТСР, оскільки помилкові позитивні результати можуть призвести до непотрібних інтервенцій. F1-Score 0,933 має високе значення, що поєднує Precision та Recall. Це свідчить про загальний баланс між точністю та здатністю моделі знаходити позитивні зразки. Щодо графіку ROC curve, він ілюструє продуктивність бінарної класифікаційної моделі шляхом порівняння показників true positive rate (TPR) та false positive rate (FPR) при різних порогах рішення [142]. ROC-криву на рис. 4.3 побудовано за бінарними рішеннями моделі, тому вона має одну точку зламу, а площа під нею 0,872 характеризує одну робочу точку і не є ранговою ROC-AUC за неперервними оцінками моделі; цим пояснюється і те, що вона нижча за AUC 0,88 моделі з квантовим шаром, ROC-криві якої побудовано за неперервними оцінками (рис. 4.1). Тому значення 0,872 у підсумкових результатах не використовується, а рангову ROC-AUC слід обчислити за ймовірностями позитивного класу (п. 4.6).

Також у рамках експерименту було досліджено можливість візуальної інтерпретації результатів, що дає навчена нейронна мережа, для кращого розуміння процесу класифікації. Для пояснюваності отриманих нейромережових рішень використано «Transformers Interpret» [143], інструмент пояснення моделі, призначений для роботи виключно з пакетом «transformers». За замовчуванням бібліотека обчислює внески токенів методом Layer Integrated Gradients.

Для контексту: у схожих за метою дослідженнях на інших даних повідомлялось про збалансовану точність 0,804 на інтерв'ю AVEC-19 [17], F1 0,81 на наративах про пологи [19], AUC 0,69 [18] і 0,74 [34] на клінічних інтерв'ю. Через відмінності наборів даних, постановок задач і протоколів оцінювання ці значення не є безпосередньо зіставними з отриманими і наведені лише як орієнтир.

Нецільовий клас навчального набору містив дописи тематично суміжних спільнот (зокрема anxiety і stress), тому модель навчалася відмежовувати ознаки ПТСР від близьких за тематикою текстів;

специфічність на валідаційній вибірці становить 0,95 (209 з 220, рис. 4.4). Окремо за сабредитами нецільового класу частоту хибних спрацьовувань не оцінювали (п. 4.6).

Кількість епох (3) обмежена обчислювальними ресурсами: навчання зайняло в сукупності 10 годин при обмеженні сеансу в Google Colab 12 годин. Монотонне зменшення втрат (loss) на навчальній вибірці свідчить лише про збіжність навчання, а достатність навченості підтверджують метрики на валідаційній вибірці.

Матриця сплутувань для валідаційного набору наведена на рис. 4.4.

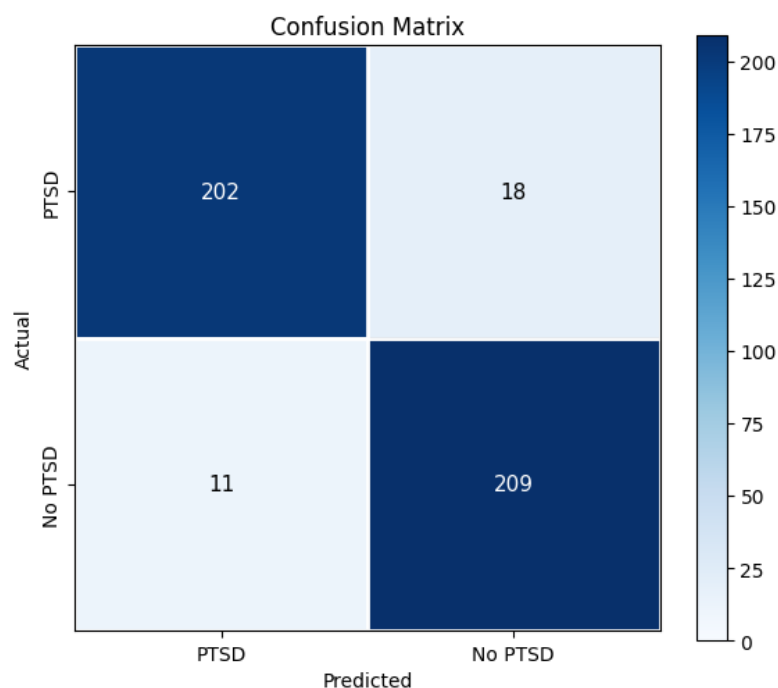


Рис. 4.4 – Матриця сплутувань

Як видно з рис. 4.4, лише 18 записів класу ПТСР з 220 ідентифіковано як відсутність прояву, а 11 записів з 220, у яких відсутній ПТСР, хибно ідентифіковано як ПТСР, що відповідає чутливості 0,918 і специфічності 0,950.

Розроблений метод дозволяє не лише ефективно виконувати аналіз прояву ПТСР в користувацькому контенті, а й надавати пояснення прийнятих

рішень. Приклад інтерпретації прийняття моделлю рішення про наявність ПТСР у тексті наведено на рис. 4.5.

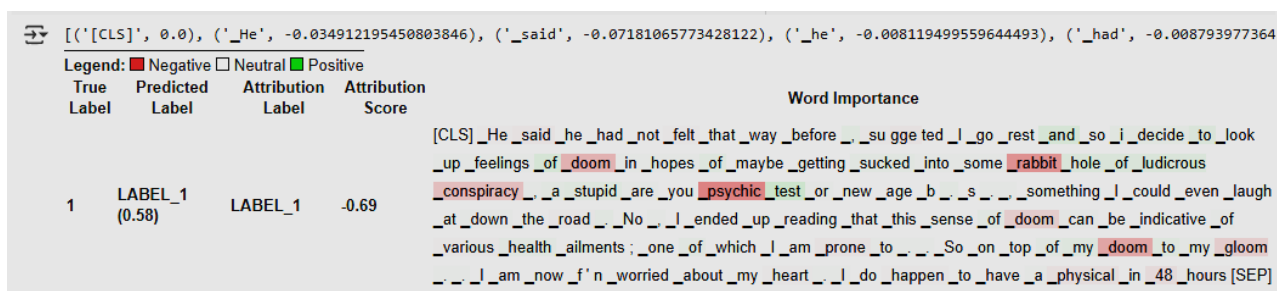


Рис. 4.5 – Інтерпретація рішення моделі-трансформера за допомогою інструменту «Transformers Interpret»

Відповідно до рис. 4.5, на вхід неймережі було подано текст: «*He said he had not felt that way before, suggeted I go rest and so i decide to look up feelings of doom in hopes of maybe getting sucked into some rabbit hole of ludicrous conspiracy, a stupid are you psychic test or new age b.s., something I could even laugh at down the road. No, I ended up reading that this sense of doom can be indicative of various health ailments; one of which I am prone to.. So on top of my doom to my gloom..I am now f'n worried about my heart. I do happen to have a physical in 48 hours*». Модель віднесла цей текст до класу з ознаками ПТСР. За атрибуцією на рис. 4.5 найбільший внесок у рішення мають слова «doom» і «gloom», пов'язані з відчуттям приреченості; помітну вагу мають і слова «conspiracy», «psychic» та «rabbit», які змістовно з ПТСР не пов'язані. Отже, пояснення дає змогу не лише побачити підстави рішення, а й виявити лексичні підказки, на які модель спирається без змістовних підстав.

В рамках виконаного дослідження також було проаналізовано як модель обробляє та інтерпретує різні частини тексту. Це особливо корисно для аналізу та вдосконалення моделей. За допомогою інструменту «BertViz» [144] візуалізація показує ваги уваги для всіх шарів і всіх головок у кожному шарі. Приклад ілюстрації такого механізму наведено на рис. 4.6, де кольором відображаються ваги уваги: чим яскравіший колір, тим більша вага уваги між

відповідними токенами; ваги уваги показують, як модель пов'язує токени між собою, але самі по собі не є мірою важливості токена для рішення.

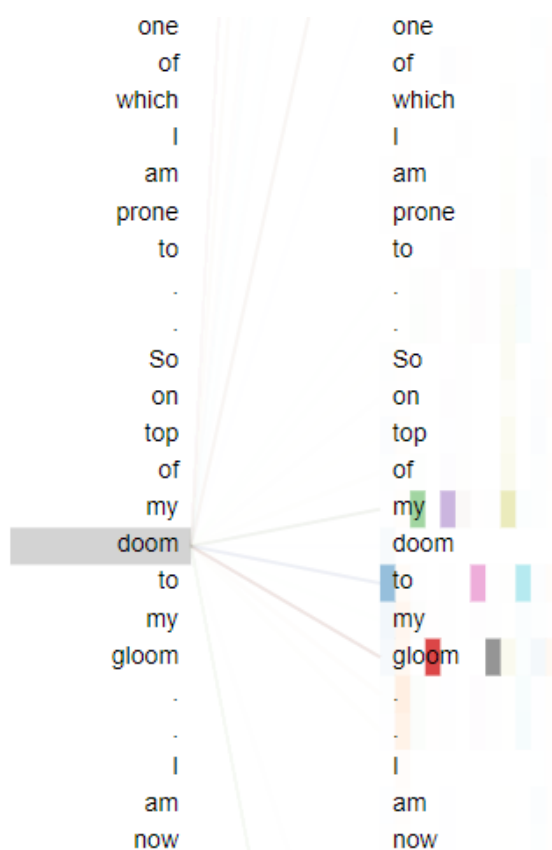


Рис. 4.6 – Інтерпретація рішення моделі-трансформера за допомогою інструменту «BertViz»

Кожен шар має кілька головок уваги, і кожна головка може мати свою власну «картину» уваги. Стрілки в візуалізації показують напрямок уваги від одного токена до іншого. Напрямок уваги вказує, який токен «дивиться» на який інший токен, а товщина і колір стрілки вказують на силу цієї уваги.

Отож, за допомогою інструментів «Transformers Interpret» та «BertViz» було одержано можливість не лише визначати наявність ПТСР у текстовому контенті, а й інтерпретувати отримані результати. Пояснюваність продемонстровано на окремих прикладах; кількісну оцінку правдивості та стабільності пояснень не виконували (п. 4.6).

4.3. Дослідження методу виявлення психічних розладів за аналізом користувацького контенту

4.3.1. Дослідження методу виявлення психічних розладів за аналізом користувацького контенту без адаптивного порогу

Щодо отриманих результатів, при багатокласовій класифікації вдалось отримати такі результати за макро-метриками: Macro Accuracy: 0,854, Macro Precision: 0,867, Macro Recall: 0,872, Macro F1-score: 0,862 (макросередні за табл. 4.2; значення 0,854, наведене раніше, є зваженим середнім і збігається з Accuracy). Матриця плутанини наведена на рис. 4.7.

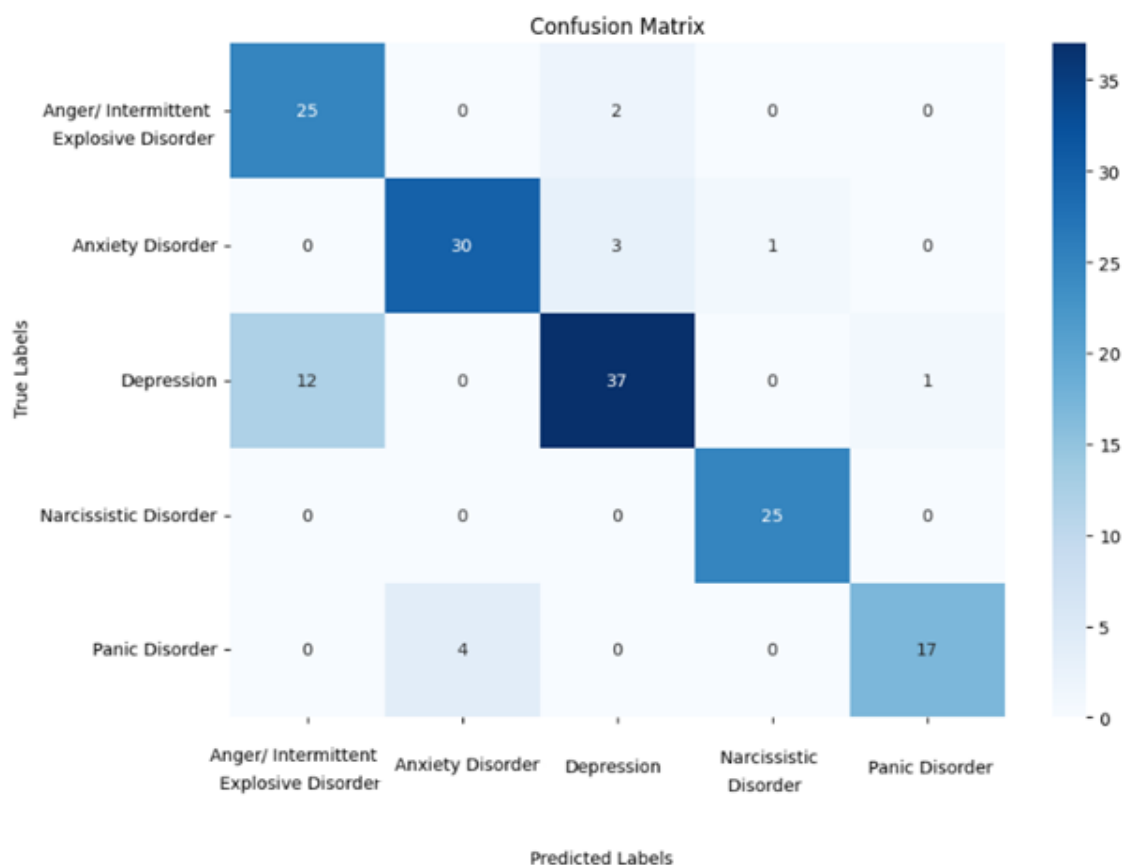


Рис. 4.7 – Матриця сплутувань для багатокласової класифікації

Що стосується метрик для кожного класу, що характеризує наявність психічних розладів, то їх значення наведені в таблиці 4.2. Кількість зразків для аналізу – 157.

Таблиця 4.2

Метрики one-vs-all при багатокласовій класифікації

Класи/Метрики	Precision	Recall	F1-score
«Anger/Intermittent Explosive Disorder»	0,68	0,93	0,78
«Anxiety Disorder»	0,88	0,88	0,88
«Depression»	0,88	0,74	0,80
«Narcissistic Disorder»	0,96	1,00	0,98
«Panic Disorder»	0,94	0,81	0,87

Згідно з даними таблиці 4.2, модель демонструє найвищі показники для класифікації «Narcissistic Disorder» (Precision – 0,96, Recall – 1,00, F1-score – 0,98), тоді як класи «Anger/Intermittent Explosive Disorder» та «Depression» мають нижчі показники F1-міри (0,78 та 0,80 відповідно). З матриці плутанини (рис. 4.7) та значень метрик видно, що класифікація розладів часто супроводжується значною кількістю помилок між класами. Це може бути обумовлено тим, що різні розлади мають спільні симптоми та прояви, а також в одному тексті можуть бути присутні декілька розладів одночасно, що суперечить підходу багатокласової класифікації.

Експеримент з багатомітковою класифікацією також не показав високих результатів, були отримані такі значення метрик: Macro Precision: 0,75, Macro Recall: 0,74, Macro F1-score: 0,75. Значення метрик є нижчими, ніж при багатокласовій. Як і у попередньому експерименті, присутня значна плутанина психічних розладів. Такі результати пов'язані з тим, що набір даних є ортогонально-розміченим, але поряд з тим деякі розлади мають схожі симптоми і можна висунути припущення, що частина текстів набору даних все ж таки попри ортогональну розмітку має тексти з декількома розладами водночас. Значення метрик наведено в таблиці 4.3, а матриця плутанини наведена на рис. 4.8.

Таблиця 4.3

Метрики one-vs-all при багатомітковій класифікації

Класи/Метрики	Precision	Recall	F1-score
«Anger/Intermittent Explosive Disorder»	0,71	0,81	0,76
«Anxiety Disorder»	0,71	0,65	0,68
«Depression»	0,76	0,74	0,75
«Narcissistic Disorder»	0,74	0,81	0,78
«Panic Disorder»	0,83	0,71	0,77

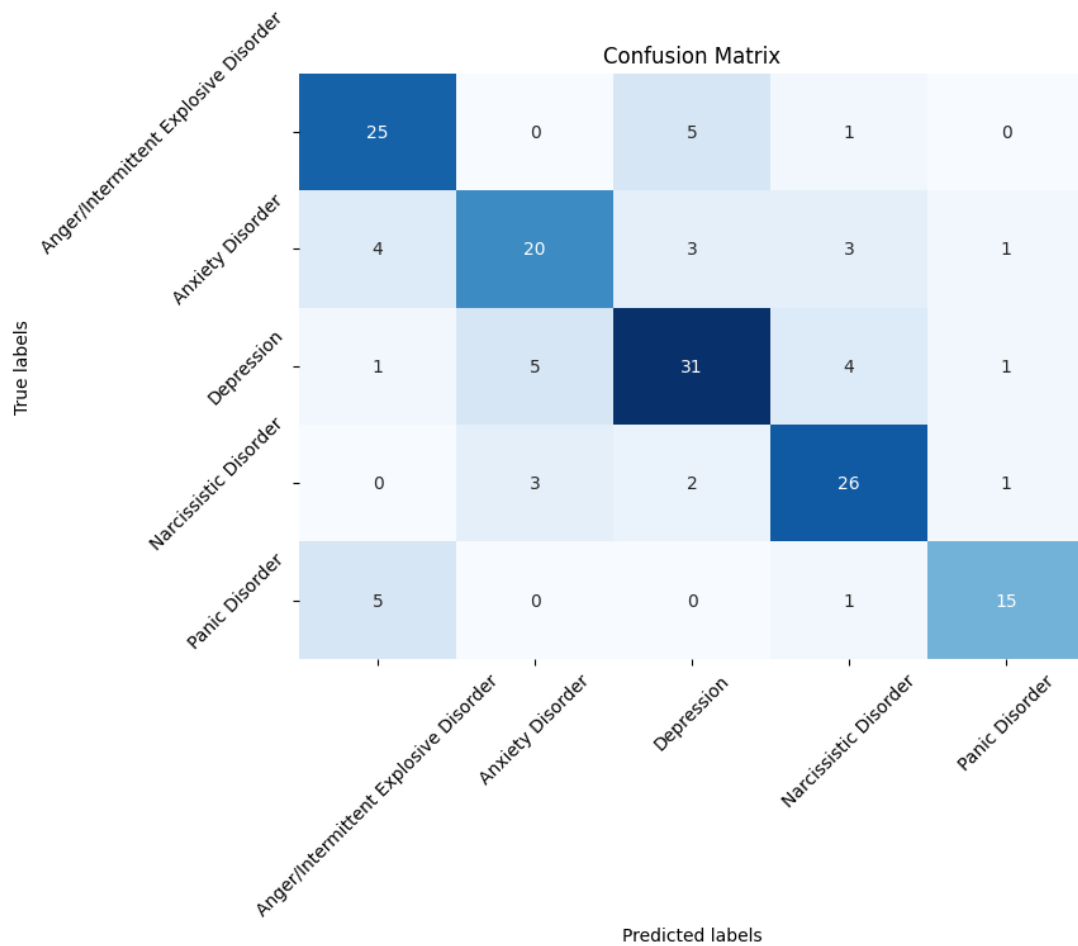


Рис. 4.8 – Матриця сплутувань для багатоміткової класифікації

Останнім експериментом було навчання класифікаторів та токенизаторів з множини M для кожного з 5 видів психічних розладів. Параметри нейромережових моделей були наступними: batch size = 16, epoch

= 5, learning_rate = $2e-5$. При такому підході були отримані результати, наведені у таблиці 4.4.

Таблиця 4.4

Метрики бінарних моделей з M

Класи/Метрики	Accuracy	Precision	Recall	F1-score	Delta F1
«Anger/Intermittent Explosive Disorder»	0,934	0,934	0,934	0,934	0,001
«Anxiety Disorder»	0,869	0,863	0,87	0,866	0,007
«Depression»	0,843	0,864	0,845	0,841	0,011
«Narcissistic Disorder»	0,984	0,986	0,982	0,984	0,003
«Panic Disorder»	0,999	0,998	0,998	0,998	0,002

У колонці $\Delta F1$ таблиці 4.4 містяться значення максимальних відхилень від середніх значень на 4-х фолдах. Precision, Recall і F1-score у табл. 4.4, 4.7 і 4.8 є макросередніми за двома класами (цільовим і нецільовим).

Матриці сплутувань бінарних класифікаторів з множини M на тестових даних наведені рис. 4.9 (a) – 4.9 (e).

Для класів «Anger/Intermittent Explosive Disorder» та «Narcissistic Disorder» модель показала високі значення метрик, що свідчить про можливість чіткої ідентифікації даних видів психічних розладів.

Для класів «Anxiety Disorder» та «Depression» спостерігаються дещо нижчі значення метрик у порівнянні з іншими класами. Хоча модель показала задовільні результати, можливе деяке змішування з іншими класами, що може бути зумовлено подібністю симптомів. Клас «Panic Disorder» має найвищі значення метрик (0,998–0,999), що підтверджується матрицею сплутувань наведеною на рис. 4.9 (b).

Щодо помилок, отриманих у «Anxiety Disorder» та «Depression» (рис. 4.9 (c) та рис. 4.9 (d)), то вони є не критичними, оскільки некоректно ідентифікуються нецільові класи.

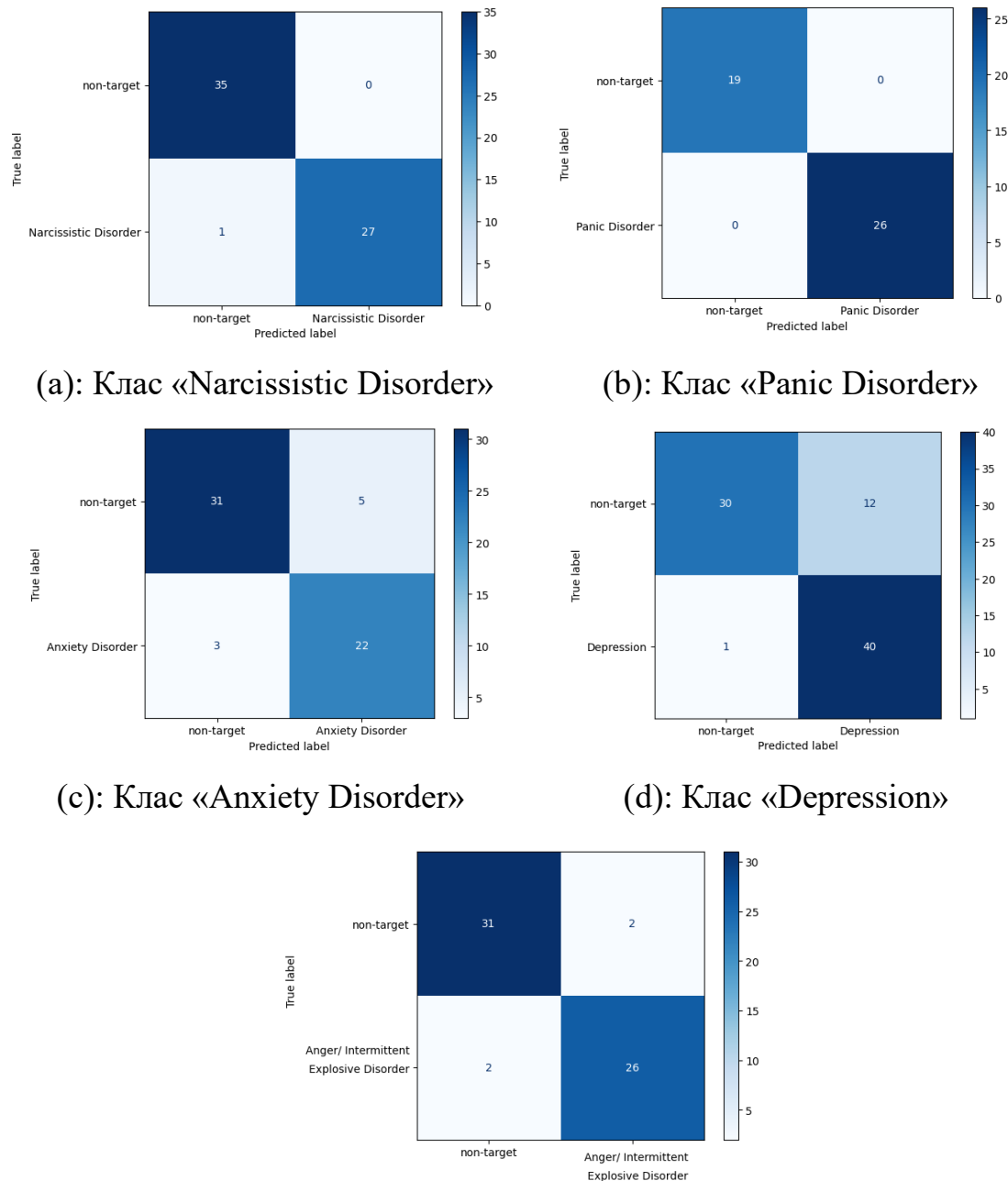


Рис. 4.9 – Матриці сплутувань бінарних класифікаторів

На рис. 4.10 наведено порівняння підходу багатокласової класифікації та бінарної для метрик Precision, Recall та F1-score.

З рис. 4.10 та таблиць 4.2, 4.4 видно, що здебільшого бінарний підхід має перевагу над багатокласовим. За допомогою використання бінарного підходу вдалось підвищити ідентифікацію «Anger/Intermittent Explosive Disorder» за метрикою Precision на 25,4 в. п., за метрикою Recall на 0,4 в. п.,

за метрикою F1-score на 15,4 в. п. Для «Depression» Precision знизилась на 1,6 в. п., Recall зросла на 10,5 в. п., а F1-score – на 4,1 в. п. Для «Narcissistic Disorder» Precision зросла на 2,6 в. п., F1-score – на 0,4 в. п., а Recall знизилась на 1,8 в. п. Для «Panic Disorder» Precision зросла на 5,8 в. п., Recall – на 18,8 в. п., а F1-score – на 12,8 в. п. Для «Anxiety Disorder» покращення досягти не вдалось. Слід враховувати, що багатокласову модель оцінено на спільній тестовій вибірці зі 157 текстів, а кожен бінарну модель – на власній збалансованій тестовій вибірці (45–83 тексти), тому порівняння на рис. 4.10 і в табл. 4.5 є орієнтовним.

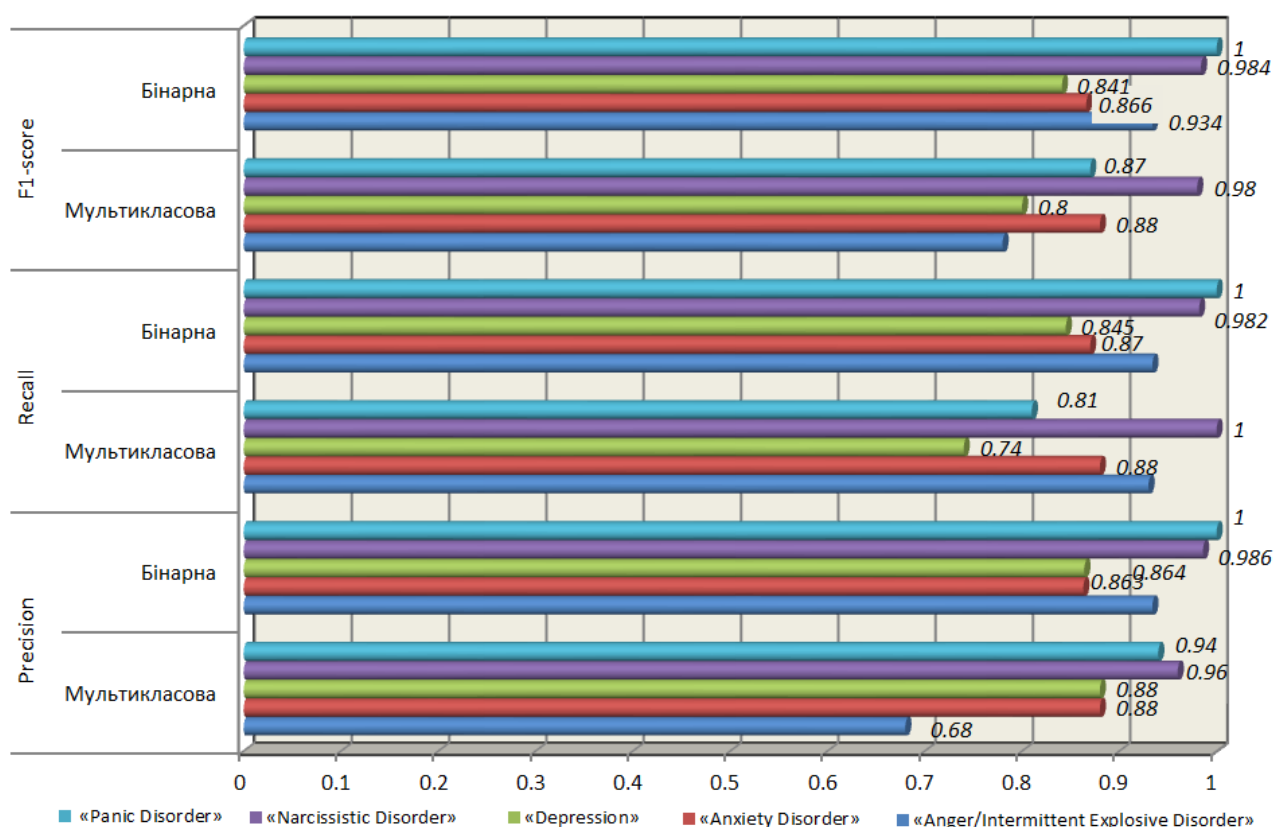


Рис. 4.10 – Порівняння підходів бінарної та багатокласової класифікації

У таблиці 4.5 наведено порівняння реалізованих підходів в межах проведеного дослідження та існуючих аналогів. Відповідно до таблиці 4.4 та таблиці 4.5, при підході застосування моделей з множини M , Macro Accuracy: 0,926, Macro Precision: 0,929, Macro Recall: 0,926, Macro F1-score: 0,925. Для контексту: у роботі [42] для багатокласової класифікації восьми класів

отримано F1 близько 0,645, у роботі [43] – 0,8965; у цьому дослідженні багатокласова модель має Macro F1-score 0,862, а множина бінарних класифікаторів – 0,925. Результати [42] і [43] отримано на інших корпусах з іншим складом класів, тому вони наведені як орієнтир, а не як доказ переваги.

Таблиця 4.5

Значення макро F1-score для альтернативних підходів

Моделі	F1-score
Багатокласовий підхід	0,862
Багатомітковий підхід	0,75
Підхід на основі множини бінарних класифікаторів	0,925
Підхід 1 [42]	0,645
Підхід 2 [43]	0,8965

Отже, запропонований підхід сприяє підвищенню коректного виявлення психічних розладів, та архітектурно дає змогу виявляти в межах одного користувацького тексту не лише найбільш проявлений розлад, а й інші наявні психічні розлади при умові що розлад вважається наявним.

4.3.2. Дослідження методу виявлення психічних розладів за аналізом користувацького контенту з адаптивним порогом

Було використано та навчено множину M для кожного з 5 видів досліджуваних психічних розладів з множини D . Обґрунтування вибору DistilBERT наведено в п. 3.3. Також використано Індекс Юдена для мінімізації ризику помилкових віднесень до категорій психічного розладу чи помилкових пропусків.

Відповідно до методики експерименту (п. 3.7), за допомогою Алгоритму 3.4 для кожної моделі з множини M було сформовано відповідний набір даних з множини наборів даних Df .

Після застосування Алгоритму 3.4, отримано такий розподіл даних в наборах даних: Df_1 містить 154 записи «Anger/ Intermittent Explosive Disorder» та 150 записів категорії «non-target»; Df_2 містить 153 записи категорії «Anxiety Disorder» та 150 записів категорії «non-target»; Df_3 містить 208 записів категорії «Depression» та 205 записів категорії «non-target»; Df_4 містить 158 записів категорії «Narcissistic Disorder» та 155 записів категорії «non-target»; Df_5 містить 112 записів категорії «Panic Disorder» 110 записів категорії «non-target».

Після виконання кроку 1 методики експерименту (п. 3.7), було отримано такі розподіли навчальних та тестових даних: для d_1 кількість зразків у тренувальному наборі: 243 (80 %), кількість зразків у тестовому наборі: 61 (20 %); для d_2 кількість зразків у тренувальному наборі: 242 (80 %), кількість зразків у тестовому наборі: 61 (20 %); для d_3 кількість зразків у тренувальному наборі: 330 (80 %), кількість зразків у тестовому наборі: 83 (20 %); для d_4 кількість зразків у тренувальному наборі: 250 (80 %), кількість зразків у тестовому наборі: 63 (20 %); для d_5 кількість зразків у тренувальному наборі: 177 (80 %), кількість зразків у тестовому наборі: 45 (20 %).

На кроці 2 експериментально було навчено множину моделей M для кожного з 5 видів психічних розладів. Параметри нейромережових моделей були наступними: batch size = 4, epoch = 4, learning_rate = $2e-5$.

Після виконання кроку 2, на кроці 3 було отримано значення функції втрат для кожної моделі з множини M (таблиця 4.6).

Значення трьох функцій втрат на навчальній і тестовій вибірках наведено в табл. 4.6. Ці функції мають різні шкали (за однакових прогнозів focal loss не перевищує log loss), тому їх абсолютні значення між собою не порівнюються і висновок про перевагу тієї чи іншої функції з табл. 4.6 не робиться; інформативним є співвідношення навчальних і тестових втрат для кожної моделі. Для «Anxiety Disorder» тестові значення focal loss і dice loss

менші за навчальні, що за малого розміру тестової вибірки (61 запис) потребує перевірки на більшій вибірці.

Таблиця 4.6

Значення функцій втрат для бінарних моделей з M

Класи / Метрики	focal loss (навчаль- на)	focal loss (тестова)	dice loss (навчальна)	dice loss (тестова)	log loss (навчальна)	log loss (тестова)
«Anger / Intermittent Explosive Disorder»	0,019	0,068	0,024	0,087	0,103	0,301
«Anxiety Disorder»	0,157	0,035	0,059	0,012	0,207	0,324
«Depression»	0,028	0,088	0,048	0,118	0,190	0,380
«Narcissistic Disorder»	0,000	0,000	0,000	0,000	0,005	0,009
«Panic Disorder»	0,009	0,002	0,023	0,019	0,068	0,080

Особливу увагу заслуговує клас «Narcissistic Disorder», для якого всі функції втрат досягли майже нульових значень як на навчальній, так і на тестовій вибірках. Це свідчить або про надзвичайно виражену відмінність цього класу від «non-target», або про низьку внутрішньокласову варіативність. Водночас у класах «Anger/Intermittent Explosive Disorder» і «Depression» зафіксовано розрив між тренувальними і тестовими втратами, що може свідчити про складність ознак, притаманних цим станам.

На кроці 4 було отримано метрики бінарних моделей з M на навчальних даних (таблиця 4.7).

Таблиця 4.7

Метрики бінарних моделей з M на навчальних даних

Класи/Метрики	Accuracy	Precision	Recall	F1-score	Delta F1
«Anger/Intermittent Explosive Disorder»	0,975	0,975	0,975	0,975	0,08
«Anxiety Disorder»	0,934	0,941	0,934	0,933	0,012
«Depression»	0,94	0,95	0,94	0,94	0,015
«Narcissistic Disorder»	1,00	1,00	1,00	1,00	0
«Panic Disorder»	0,977	0,978	0,977	0,977	0,011

У колонці $\Delta F1$ таблиці 4.7 містяться значення максимальних відхилень від середніх значень на 4-х фолдах для навчальних даних. Отримані метрики свідчать про ефективність навчання бінарних моделей для кожної з цільових категорій психічних розладів. Усі моделі продемонстрували значення F1-міри вище 0,93, що підтверджує їх здатність до точного відтворення патернів у навчальних вибірках. Найвищу узгодженість між метриками продемонстрували моделі для «Anxiety Disorder» та «Panic Disorder», де спостерігається мінімальна різниця між показниками точності, повноти та F1. Водночас найбільше значення $\Delta F1$ зафіксовано для моделі «Anger/Intermittent Explosive Disorder», що, у поєднанні з тестовими втратами, може свідчити про перенавчання. Повна відсутність помилки у випадку «Narcissistic Disorder» потребує додаткового тестування, оскільки може вказувати як на виняткову роздільність класу, так і на потенційне перенавчання. Загалом моделі демонструють спроможність до класифікації в умовах збалансованих даних.

У межах експериментального оцінювання метрика Hamming Loss не використовувалася, оскільки застосовані набори даних не містять вичерпної мультиміткової розмітки для кожного тексту. У зв'язку з цим її використання могло б призвести до некоректної інтерпретації якості класифікації. Натомість оцінювання здійснювалося за метриками Accuracy, Precision,

Recall, F1-score та ROC-AUC, які є коректними для аналізу незалежних бінарних класифікаторів у складі запропонованого підходу.

Після аналізу навчальних даних, на кроці 5 відбувається аналіз метрик на тестових даних (таблиця 4.8).

Таблиця 4.8

Метрики бінарних моделей з M на тестових даних

Класи/Метрики	Accuracy	Precision	Recall	F1-score	Delta F1
«Anger/Intermittent Explosive Disorder»	0,92	0,92	0,92	0,92	0,03
«Anxiety Disorder»	0,89	0,88	0,88	0,88	0,049
«Depression»	0,87	0,89	0,87	0,87	0,06
«Narcissistic Disorder»	1,00	1,00	1,00	1,00	0
«Panic Disorder»	0,978	0,979	0,978	0,978	0,02

У колонці $\Delta F1$ таблиці 4.8 містяться значення максимальних відхилень від середніх значень на 4-х фолдах. Серед бінарних моделей найвищі метрики на тестовій вибірці зафіксовано для категорії «Narcissistic Disorder» (усі значення дорівнюють 1,00, $\Delta F1 = 0$), що вказує на повну класифікаційну точність. Водночас, результати можуть бути зумовлені компактним розміром набору даних та простотою розділення цього класу. Стабільність також демонструє модель для «Panic Disorder» ($F1 = 0,978$, $\Delta F1 = 0,02$), яка, на відміну від попередньої, показує узгоджену поведінку без ознак надмірної адаптації. Помірно нижчі значення F1-міри для моделей «Anxiety Disorder» та «Depression» (відповідно 0,88 і 0,87) супроводжуються вищим розкидом між фолдами ($\Delta F1 = 0,049$ та 0,06), що вказує на складнішу класифікаційну задачу в межах цих категорій. У цілому, усі моделі демонструють узагальнювальну здатність, проте окремі результати вимагають обережної інтерпретації з урахуванням кількісних та якісних характеристик даних.

Додатково нижче наведено матриці помилок для кожної із класифікаційних моделей множини M та ROC-криві. Для «Anxiety Disorder» оптимальний поріг (за індексом Юдена): 0,7401 матриця помилок та ROC-крива наведені на рис. 4.11

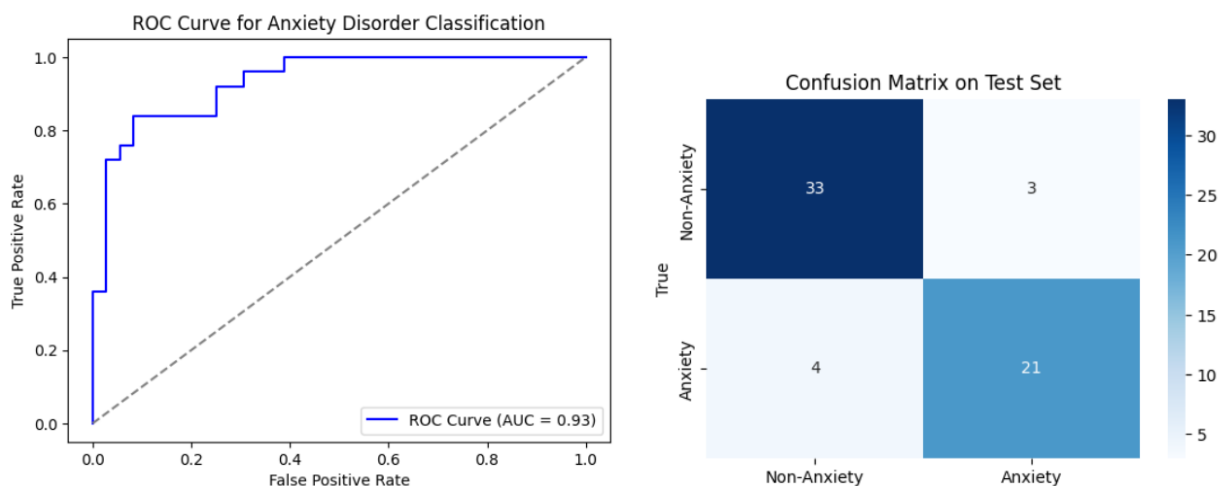


Рис. 4.11 – Графіки класу Anxiety Disorder

Для «Anger/Intermittent Explosive Disorder» оптимальний поріг (за індексом Юдена): 0,6288. Матриця помилок та ROC-крива наведені на рисунку рис. 4.12.

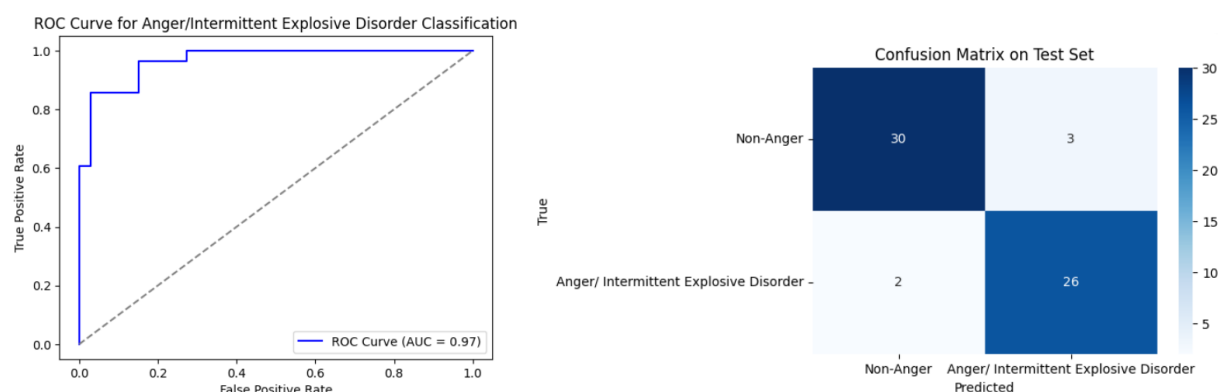


Рис. 4.12 – Графіки класу «Anger/Intermittent Explosive Disorder»

Для «Depression» оптимальний поріг (за індексом Юдена): 0,501. Матриця помилок та ROC-крива наведені на рис. 4.13.

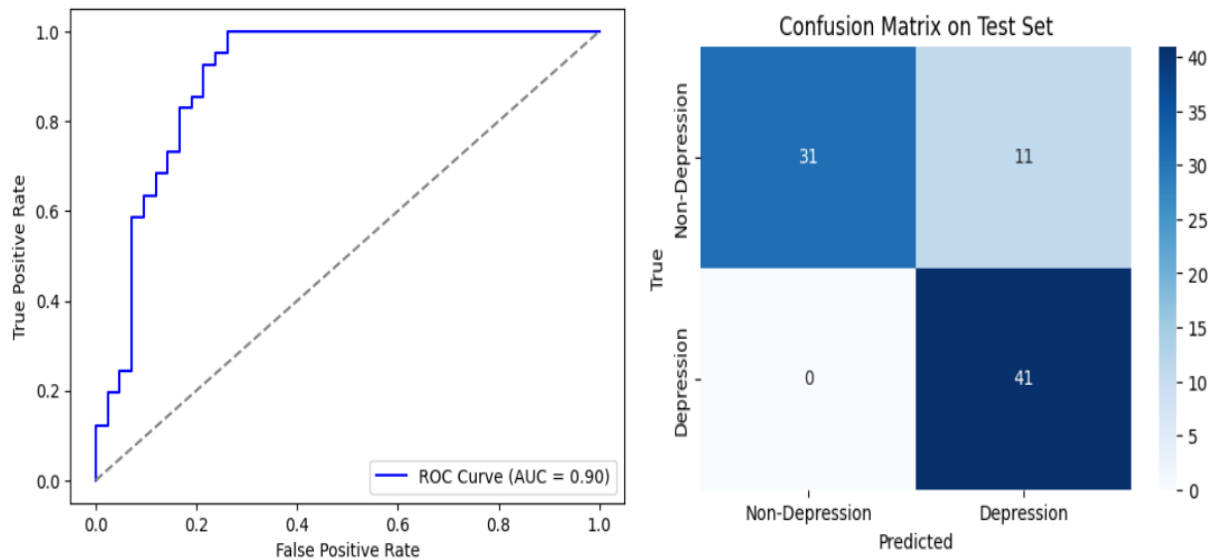


Рис. 4.13 – Графіки класу «Depression»

Для «Panic Disorder» оптимальний поріг (за індексом Юдена): 0,6094. Матриця помилок та ROC-крива наведені на рис. 4.14.

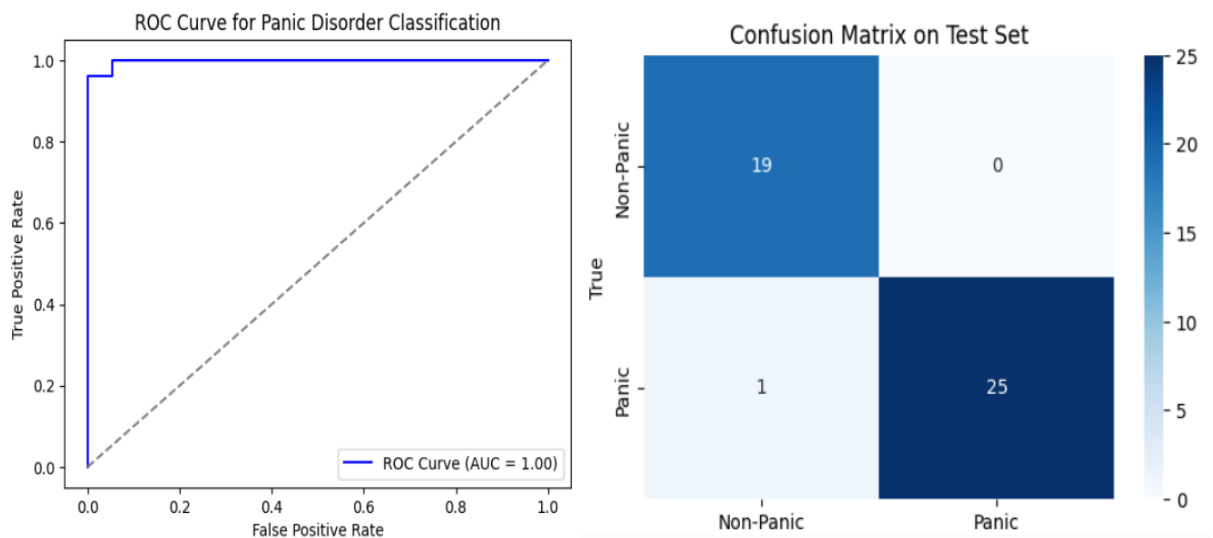


Рис. 4.14 – Графіки класу «Panic Disorder»

Для «Narcissistic Disorder» оптимальний поріг (за індексом Юдена): 0,8978. Матриця помилок та ROC-крива наведені на рис. 4.15.

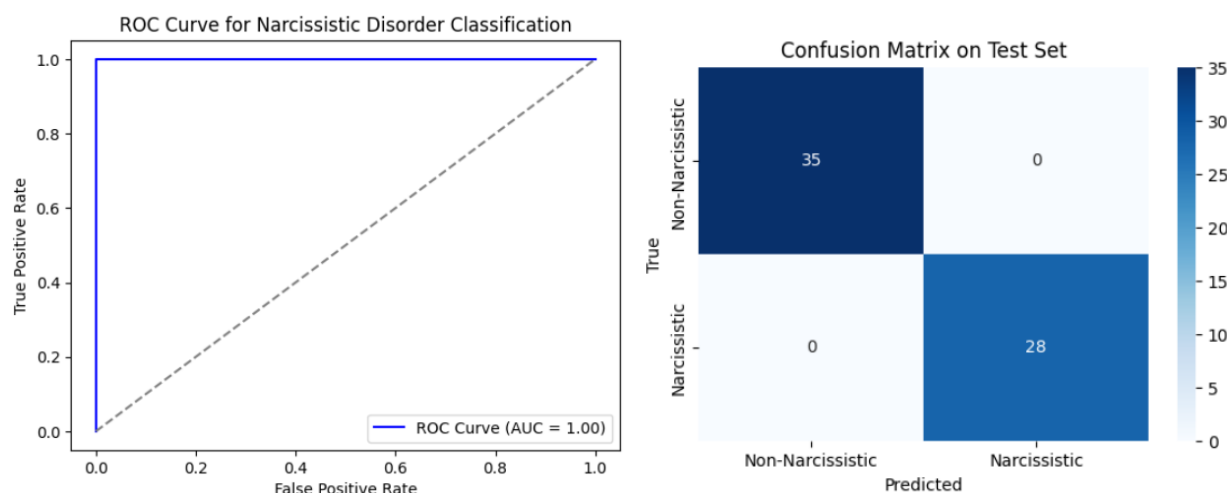


Рис. 4.15 – Графіки класу «Narcissistic Disorder»

Оптимальні пороги класифікації, визначені за індексом Юдена, та відповідні матриці помилок і ROC-криві (рис. 4.11–4.15) дають змогу оцінити розрізняльну здатність моделей для кожного класу. Для «Anxiety Disorder» за порога 0,7401 отримано 33 істинно негативних, 21 істинно позитивне, 4 хибнопозитивних і 3 хибнонегативних рішення. Для «Anger/Intermittent Explosive Disorder» за порога 0,6288 – 30 істинно негативних, 26 істинно позитивних, 2 хибнопозитивних і 3 хибнонегативних рішення, тобто помилки обох типів поодинокі. Для «Depression» за порога 0,501 – 41 істинно позитивне рішення без хибнонегативних, але 11 хибнопозитивних, тобто висока чутливість за зниженої специфічності. Для «Panic Disorder» за порога 0,6094 – 25 істинно позитивних, 19 істинно негативних, одне хибнопозитивне рішення і жодного хибнонегативного. Для «Narcissistic Disorder» за порога 0,8978 помилок немає (28 істинно позитивних і 35 істинно негативних рішень).

Усі моделі мають AUC у межах 0,89–1,00. Відмінності порогів (від 0,501 до 0,8978) показують, що оцінки різних моделей не є зіставними між собою як ймовірності, тому кожній моделі потрібен власний поріг. Для моделей «Depression», «Panic Disorder» і «Narcissistic Disorder» характерна висока чутливість за меншої специфічності, тобто схильність до

хибнопозитивних рішень. Тестові вибірки малі (від 45 до 83 записів), тому наведені значення мають широкі довірчі інтервали (п. 4.6).

Кроки 6 та 7 враховані у поданні результатів кроків 1–5. Нижче наведено порівняння отриманих результатів з відомими аналогами. У таблиці 4.9 наведено порівняння реалізованих підходів в межах проведеного дослідження та існуючих аналогів.

Таблиця 4.9

Значення макро F1-score для альтернативних підходів

Моделі	F1-score
Підхід на основі множини бінарних класифікаторів	0,9296
Підхід 1 [42]	0,645
Підхід 2 [43]	0,8965

Відповідно до таблиці 4.8 та таблиці 4.9, при підході застосування моделей з множини *M*, Macro Accuracy: 0,9316, Macro Precision: 0,9338, Macro Recall: 0,9296, Macro F1-score: 0,9296. Як і в попередньому порівнянні, результати робіт [42] (F1 близько 0,645) і [43] (0,8965) отримано на інших корпусах, тому їх наведено лише як орієнтир.

Отже, запропонований підхід сприяє підвищенню коректного виявлення психічних розладів, та архітектурно дає змогу виявляти в межах одного користувацького тексту не лише найбільш проявлений розлад, а й інші наявні психічні розлади при умові що розлад вважається наявним (якщо сила його прояву вища за поріг, що визначається пошуком точки на ROC-кривій, яка відповідає найкращому співвідношенню чутливості та специфічності). Оскільки використані набори мають розмітку з однією міткою на текст, експеримент підтверджує коректність бінарних моделей, а не здатність виявляти реальну коморбідність, перевірка якої потребує повністю багатоміткового набору даних.

4.4. Дослідження методу розширення множини ознак користувачького контенту для попередження соціально небезпечної поведінки

Нейромережеві моделі, використані для виявлення ознак ПТСР та супутніх психічних розладів у текстах, були запозичені з власних попередніх досліджень. Для моделі, що застосовується для ідентифікації ПТСР-контенту, у процесі навчання на валідаційних даних було отримано такі показники: Accuracy – 0,934, Precision – 0,948 та F1-score – 0,933 [158]. Для супутніх психічних розладів використано 5 незалежних моделей, які показали такі результати [159]: для «Anger/Intermittent Explosive Disorder» – Accuracy 0,934; для «Anxiety Disorder» – 0,869; для «Depression» – 0,843; для «Narcissistic Disorder» – 0,984; а для «Panic Disorder» – 0,999 (конфігурація без адаптивних порогів, табл. 4.4).

У цьому експерименті виходи першого та другого методів не переописуються, а використовуються як уже валідовані складові розширеного профілю. Емоційний контекст повідомлень визначається моделлю GoEmotions [131], після чого аналіз виконується у часовому вимірі відповідно до (2.16)–(2.18).

Для визначення мітки тональності використано модель [131], навчену на корпусі «GoEmotions», що охоплює 27 категорій емоційних станів і нейтральний клас, що забезпечує узагальнювальну здатність при аналізі текстів із вираженим емоційним забарвленням. У межах даного дослідження отримані емоційні стани було згруповано в узагальнені категорії «негативні» та «не негативні», що дозволило підвищити інтерпретованість результатів і спростити подальший аналіз.

Відповідно, виконавши аналіз усіх повідомлень набору даних на ознаки наявності ПТСР було отримано графік, наведений на рис. 4.16.

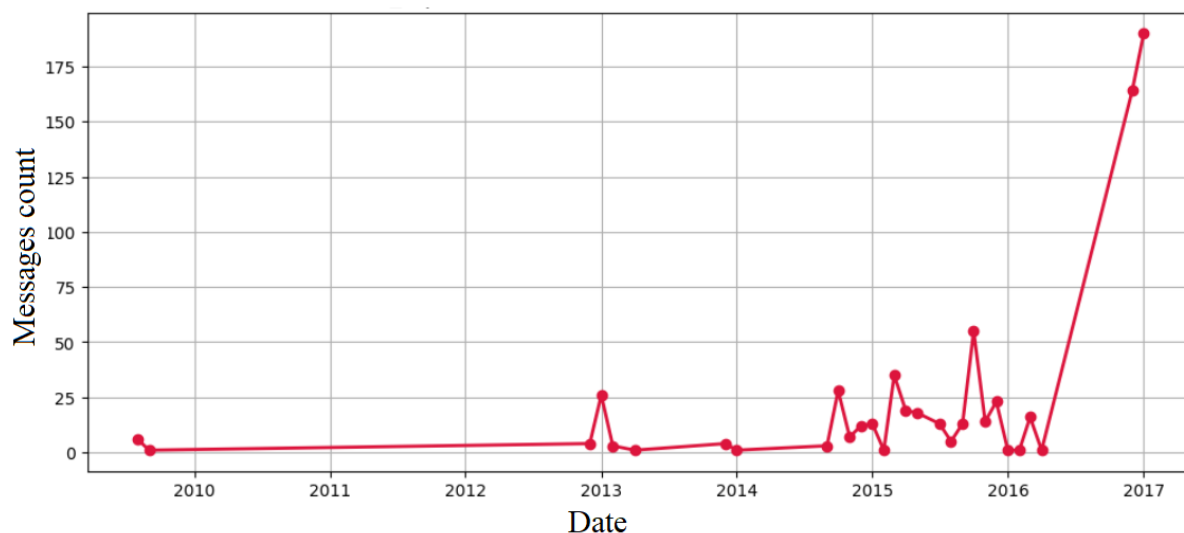


Рис. 4.16 – Графік аналізу повідомлень щодо наявності ознак ПТСР

На графіку відображено часову динаміку кількості повідомлень, що класифіковані як такі, що містять ознаки ПТСР. По осі абсцис показано часову шкалу, яка охоплює період з 2009 до початку 2017 року, а по осі ординат – кількість відповідних повідомлень. Візуалізація демонструє, що протягом тривалого часу інтенсивність появи таких постів залишалася мінімальною або майже відсутньою. Лише з 2013 року почали фіксуватися окремі сплески активності, проте вони мали епізодичний характер і не формували стійкої тенденції. Починаючи з 2015 року, динаміка стає більш вираженою: спостерігається поступове зростання кількості повідомлень, яке супроводжується нерівномірними коливаннями. Найбільш різке збільшення кількості таких повідомлень зафіксовано наприкінці досліджуваного періоду – у 2016 – на початку 2017 року, коли відбувається стрімкий стрибок частоти виявлених ПТСР-постів. Оскільки графік відображає абсолютну кількість виявлених повідомлень без нормування на загальний обсяг дописів у кожному періоді, зростання наприкінці періоду може бути зумовлене нерівномірністю обсягу даних у наборі, тому змістовних висновків щодо поширеності ПТСР-дискурсу з нього не робиться.

У наявному наборі даних останні повідомлення датуються 2017 роком. Для змістовної інтерпретації часової динаміки в подальших дослідженнях

доцільно аналізувати частку ПТСР-повідомлень у загальній кількості дописів за період.

Результати аналізу ризикової поведінки за користувачами наведено на рис. 4.17–4.19.

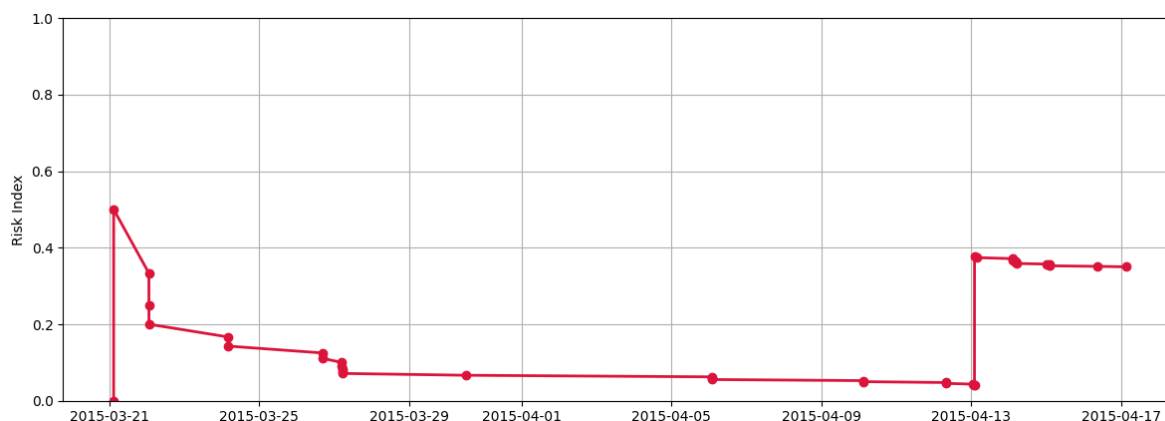


Рис. 4.17 – Графік аналізу часового ряду щодо можливої ризикової поведінки зі сплесками

Поданий графік відображає динаміку індексу ризику користувача на основі тридцятиденного ковзного вікна. У часовому розрізі спостерігається кілька характерних фаз: початковий різкий сплеск значень індексу, поступове його зниження до майже мінімальних рівнів, а згодом – новий етап стабілізації на підвищеному рівні. Така траєкторія значною мірою зумовлена властивостями формули (2.16): на початку спостереження вікно містить мало повідомлень, тому одне повідомлення з ознаками ПТСР дає різкий сплеск, після якого з надходженням нових повідомлень індекс спадає; стрибок наприкінці ряду відповідає спрацюванню порогових індикаторів (2.17). Тому такі зміни не слід тлумачити як кризові періоди чи ремісію.

Наприклад, перше повідомлення цього користувача – саркастичний допис із ненормативною лексикою, у якому автор заперечує «факти» і жартома називає себе єдинорогом, – модель рівня повідомлення визначила як таке, що містить ознаки ПТСР, а моделі супутніх розладів – як депресивне. Змістовно це саркастичний допис без клінічних проявів, тобто

хибнопозитивне рішення методів рівня повідомлення. Його причина пояснювана: ненормативна лексика та агресивні дієслівні конструкції є частотними маркерами цільових класів у навчальних вибірках, тоді як іронія та сарказм у них не розмічені. Показово, що в наборі даних цей допис також має мітку депресивного стану, тобто й розмітка джерела не відрізняє сарказм від клінічного прояву; тому узгодженість індексу з розміткою слід трактувати на рівні користувача та часового вікна, а не окремого допису. За (2.16) одне повідомлення змінює базову складову індексу не більше ніж на $1/|I_u(t, W)|$: за малого наповнення вікна (перші дописи спостереження) одиничне хибнопозитивне рішення дає помітний короткий сплеск, що видно на початку рис. 4.17, а зі збільшенням кількості повідомлень у вікні його вплив зменшується, і стійко підвищений рівень індексу наприкінці рис. 4.17 утворюється вже серією повідомлень. Тому інтерпретувати слід стійкі рівні індексу за наповненим вікном, а не поодинокі сплески на його початку. Цей випадок ілюструє відмінність між точністю на рівні повідомлення та якістю оцінки на рівні користувача; обмеження щодо сарказму зафіксовано в п. 4.6.

Також при аналізі ризикової поведінки спостерігається ряд графіків без сплесків, наприклад, як на рис. 4.18.

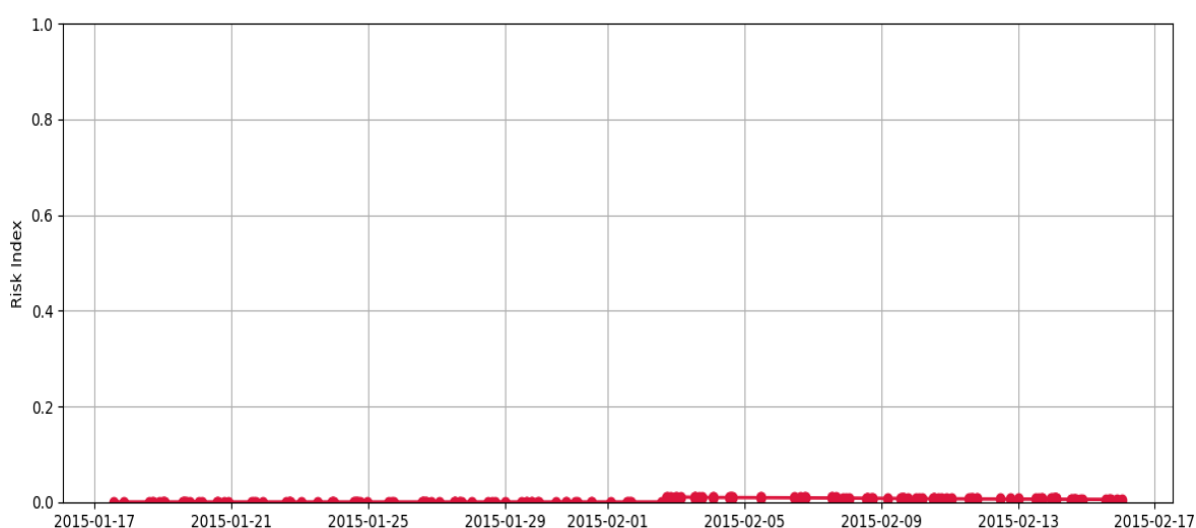


Рис. 4.18 – Графік аналізу часового ряду щодо можливої ризикової поведінки з відсутністю коливаний

Аналіз публікацій користувача, профіль якого наведено на рис. 4.18, свідчить про домінування мотиваційної, спортивної та побутової тематики без систематичних проявів негативної спрямованості чи ознак ПТСР. Єдина згадка про втрату близької людини має радше контекст подолання труднощів і самопідтримки, ніж фіксацію тривалих деструктивних станів. Це безпосередньо узгоджується з отриманим графіком, на якому індекс ризику протягом усього періоду залишається на мінімальному рівні, підтверджуючи відсутність значних коливань, що могли б свідчити про психологічну кризу чи стійкі негативні патерни у поведінці користувача.

Однак, метод поки має обмеження. Наприклад, під час аналізу графіку, що наведений на рис. 4.19, більшість постів показали високі значення індексу ризику.

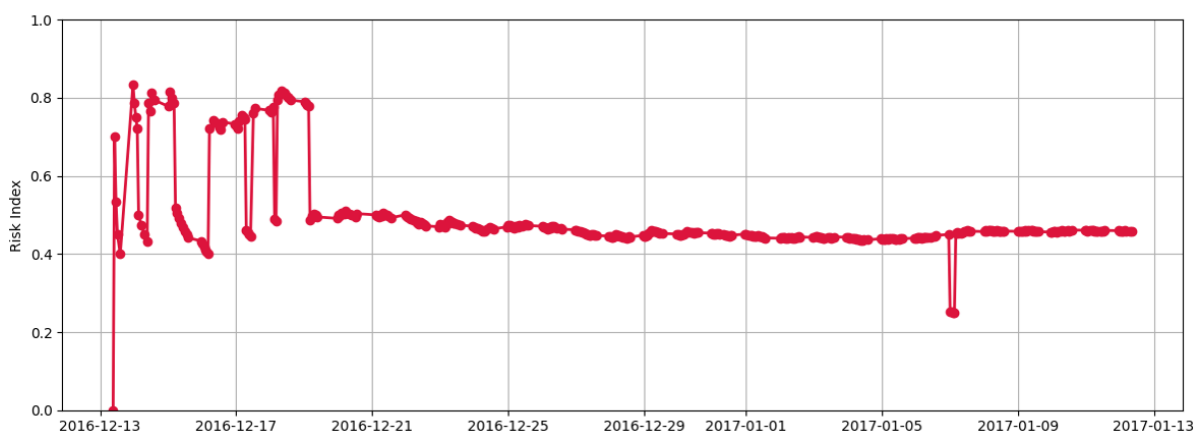


Рис. 4.19 – Графік аналізу часового ряду щодо можливої ризикової поведінки з високими індикаторами

Серед дописів траплялися жартівливі згадки про вбивство людини та про «вбивство бога і прокляття нескінченної реінкарнації», які за лексикою виглядають агресивними. Проте повний перегляд дописів автора, серед яких є репліки про «грізного супротивника» та ігрових персонажів, показує, що йдеться, найімовірніше, про ігрових персонажів і досвід у відеоіграх, а не про реальне життя. Це обумовлено навчальною вибіркою, на якій донавчалися

нейромережі для виявлення ПТСР та коморбідних розладів. Тому для ідентифікації таких постів нейромережі необхідно додатково донавчати.

Проведений аналіз засвідчує, що хоч і запропонований підхід має обмеження, однак він дає змогу виділяти профілі зі стійко підвищеним індексом ризику для подальшого перегляду фахівцем. Важливим є те, що модель враховує не лише наявність ПТСР, а й супутню коморбідність та негативне емоційне забарвлення дописів, які виступають додатковими підсилюючими факторами. Саме через ці компоненти ризиковий індекс набуває більшої чутливості та відображає складну багатовимірну картину психоемоційного стану користувача, дозволяючи інтерпретувати його динаміку не як просте накопичення епізодів, а як системний показник, що поєднує різні аспекти дисфункціональної поведінки.

Отже, проведені експерименти з аналізом часових рядів ризикової поведінки показали якісну узгодженість окремих результатів із наявною міткою депресивного стану в досліджуваному наборі даних та з ручним переглядом дописів. Це підтверджує працездатність механізму інтеграції ознак і часової агрегації, тоді як кількісна оцінка прогностичної точності індексу ризику щодо соціально небезпечної поведінки потребує окремо розміченого набору даних.

Внесок методу в досягнення мети показано наявними результатами: індекс розрізняє профілі з епізодичними сплесками (рис. 4.17), без коливань (рис. 4.18) та зі стійко високим рівнем (рис. 4.19), тоді як на рівні повідомлень ці користувачі описуються лише наборами міток; підвищені значення індексу узгоджуються з міткою депресивного стану в наборі даних та ручним переглядом дописів; завдяки декомпозиції (2.18) для профілю на рис. 4.19 встановлено складові високого індексу та виявлено ігровий контекст як джерело хибнопозитивних рішень. Клінічна прогностична точність індексу потребує окремо розміченого набору даних (п. 4.6).

4.5. Узагальнення результатів експериментального дослідження

Експерименти підтвердили різні, але взаємодоповнювальні аспекти трьох наукових результатів. Для методу виявлення ПТСР модель DeBERTa досягла Accuracy 0,934, Precision 0,948 та F1 0,933 на валідаційній вибірці; окрема модель із квантовим шаром (Accuracy 0,85) розглядається як пошуковий експеримент. Ці результати перевіряють контекстно-позиційну складову першого методу та спеціалізоване формування навчальних даних.

Для методу виявлення психічних розладів перехід від одного взаємовиключного класу до сукупності бінарних моделей усуває структурне обмеження багатокласового рішення. У варіанті з адаптивними порогами отримано Macro Accuracy 0,9316, Macro Precision 0,9338, Macro Recall 0,9296 та Macro F1-score 0,9296. Порівняння з багатокласовими й опублікованими аналогами наведено у п. 4.3 та працях [42, 43].

Третій метод перевіряється не як ще один ізольований класифікатор, а як механізм інтеграції та часової агрегації ознак. Отримані часові ряди демонструють, як спільна поява ПТСР-контенту, супутніх розладів і негативного емоційного контексту змінює індекс ризику. Тому результат цього сценарію слід трактувати як валідацію працездатності формалізованого механізму розширення ознак і відстеження динаміки, а не як доказ клінічної прогностичної точності, що окремо враховано в обмеженнях.

4.6. Обмеження запропонованих методів та використаних наборів даних

Незважаючи на отримані позитивні результати експериментального дослідження, запропоновані в дисертаційній роботі методи та їх програмна реалізація мають низку обмежень, зумовлених як властивостями використаних наборів даних, так і особливостями предметної області аналізу психічного стану людини за користувацьким контентом. Урахування таких

обмежень є важливим для коректної інтерпретації одержаних результатів, оцінювання меж застосовності розроблених методів і визначення напрямів подальшого розвитку дослідження.

Першу групу обмежень становлять обмеження, пов'язані з використаними наборами даних. У роботі застосовано відкриті текстові набори даних, сформовані на основі дописів із соціальних мереж та інших онлайн-ресурсів. Такі дані є цінним джерелом для інтелектуального аналізу психічного стану, оскільки відображають природну мовленнєву поведінку користувачів, однак вони не є повністю еквівалентними клінічно валідованим вибіркам. Тексти у відкритих наборах даних зазвичай формуються в неоднорідних комунікативних умовах, містять значний рівень стилістичної варіативності, різну довжину, різну емоційну насиченість та різний ступінь контекстної завершеності. Унаслідок цього частина ознак, виявлених моделями, може бути пов'язана не лише з проявами психічного стану, а й зі специфікою конкретної платформи, жанру онлайн-комунікації або індивідуальним стилем самовираження автора.

Окремим обмеженням є неоднорідність джерел даних для різних підзадач дослідження. Для виявлення ПТСР, виявлення інших психічних розладів і оцінювання ризиків соціально небезпечної поведінки використовуються тексти, що можуть відрізнятися за тематикою, мовною структурою, способом розмітки та принципами формування вибірки. Це ускладнює повну уніфікацію експериментальних умов і може впливати на безпосередню порівнюваність результатів між окремими методами. Крім того, використані дані відображають лише ту частину цифрової активності користувачів, яка стала доступною для збору та подальшої обробки, тоді як реальна картина психоемоційного стану людини може формуватися на основі значно ширшого контексту, включно з невербальними сигналами, історією взаємодії, часовою динамікою змін стану та зовнішніми життєвими обставинами.

Для задачі виявлення ПТСР істотним обмеженням є складність формування репрезентативного навчального набору даних, у якому ознаки ПТСР були б відмежовані від проявів інших психічних станів. Навіть за умови спеціального формування негативного класу з текстів без проявів ПТСР та текстів, що можуть належати до інших психічних розладів, залишається ризик часткового перетину мовних маркерів. Це пояснюється тим, що ПТСР, тривожні стани, депресивні прояви, емоційне виснаження або агресивні реакції можуть мати спільні лексичні, синтаксичні й емоційно-семантичні характеристики. Відповідно, навіть контекстно-орієнтована модель не завжди може однозначно відокремити специфічні прояви ПТСР від суміжних психічних станів у всіх типах текстів.

Ще одним обмеженням першого методу є залежність якості виявлення ПТСР від повноти контексту в окремому текстовому фрагменті. У реальних умовах користувацький контент часто складається з коротких дописів, уривчастих висловлювань, фрагментованих емоційних реакцій або іронічних конструкцій, зміст яких істотно залежить від попередніх повідомлень, зовнішньої ситуації чи невербального підтексту. У таких випадках навіть модель, яка враховує контекстні залежності й позиційні взаємозв'язки між словами, може працювати з неповною інформацією. Це, у свою чергу, обмежує стабільність узагальнення результатів на тексти, що суттєво відрізняються від навчальної вибірки за структурою, тематикою або способом вираження емоційних станів.

Окремим обмеженням методів рівня повідомлення є сарказм та іронія: саркастичні дописи з ненормативною чи агресивною лексикою можуть отримувати хибнопозитивні рішення щодо ПТСР і супутніх розладів (приклад у п. 4.4), оскільки такі явища не розмічені в навчальних вибірках. На рівні користувача їх вплив послаблюється усередненням у часовому вікні (2.16), (2.23), проте повне усунення потребує окремого детектора сарказму, що є напрямом подальших досліджень.

Для другого методу, спрямованого на виявлення психічних розладів за аналізом користувацького контенту, ключовим обмеженням є клінічна та лінгвістична близькість окремих класів. Особливо це стосується тривожних і депресивних проявів, які нерідко мають схожі мовні маркери: негативне емоційне тло, ознаки безнадійності, високу частоту слів із семантикою напруження, страху, втоми, самооцінних знижень або дезадаптації. Саме тому навіть за умов роздільного підходу до класифікації та побудови незалежних бінарних моделей повністю усунути помилки між клінічно близькими станами неможливо. Це обмеження є не лише технічним, а й змістовним, оскільки межі між окремими психічними станами в текстовому самовираженні людини часто є нечіткими.

Окремо слід зазначити особливість розмітки використаних даних, яка є умовою задачі, а не дефектом наборів: розмітка є повною щодо цільового класу кожного тексту, тому бінарні моделі множини M оцінюються коректно, а неповнота стосується лише перевірки одночасної наявності кількох станів. Хоча розроблений підхід орієнтований на виявлення кількох психічних розладів одночасно, застосовані набори даних не завжди містять вичерпну мультиміткову анотацію для кожного тексту. Для окремого допису може бути гарантовано наявність певного цільового класу, однак не гарантовано повне відображення всіх можливих супутніх станів. Це означає, що модель може коректно виявляти частину прихованих або супутніх ознак, які формально не відображені в розмітці, а отже, не завжди можуть бути об'єктивно оцінені в межах наявного експериментального протоколу. Саме тому результати багатоміткового аналізу слід трактувати як оцінку здатності моделі виявляти потенційні коморбідні прояви, а не як повністю вичерпне відтворення реальної психіатричної картини.

Для третього методу, пов'язаного з розширенням множини ознак користувацького контенту з метою попередження соціально небезпечної поведінки, важливим обмеженням є непрямий характер зв'язку між виявленими текстовими ознаками та реальною поведінкою людини. Такі

ознаки, як загальна тональність тексту, різка зміна емоційного забарвлення, рівень агресії, частота негативних висловлювань, наявність небезпечної лексики, кількість граматичних помилок або особливості емоційного профілю, можуть вказувати на підвищений ризик, однак самі по собі не є достатньою підставою для висновку про реальну готовність людини до соціально небезпечної поведінки. Вони мають імовірнісний характер і повинні розглядатися як допоміжні індикатори, що сигналізують про потребу в додатковому аналізі, а не як остаточний доказ небезпеки.

Крім того, частина ознак третього методу чутлива до мовної норми, стилю письма, вікових особливостей користувача, платформи спілкування та індивідуальної манери самовираження. Наприклад, велика кількість граматичних помилок може бути наслідком не психологічного напруження, а низької уваги до письма, технічних особливостей введення тексту або специфіки неформального онлайн-спілкування. Аналогічно, високий рівень емоційної експресії або використання агресивної лексики не завжди означає наявність реального ризику, оскільки може бути зумовлений ситуативним афектом, жанровими особливостями комунікації, мовною грою або стилістикою певної онлайн-спільноти. Це означає, що інтерпретація результатів третього методу повинна здійснюватися з урахуванням ширшого комунікативного контексту. Саме тому метод розширення множини ознак призначений для скринінгу та ранжування користувачів за рівнем ризику, а не для остаточного висновку щодо окремого повідомлення: рухоме вікно (2.16) згладжує одиничні сплески, зумовлені ігровим чи іронічним контекстом, і виділяє лише стійкі тенденції.

Окрему групу становлять обмеження, пов'язані з експериментальним дизайном та статистичною стійкістю результатів. Для частини класів використовуються відносно компактні вибірки, що може знижувати стійкість оцінок при різних випадкових розбиттях на навчальні та тестові підвибірки. Для зменшення впливу цього чинника метрики другого методу усереднено за чотирма випадковими розбиттями з наведенням максимальних відхилень $\Delta F1$

(табл. 4.4, 4.7, 4.8), що дає змогу оцінити стійкість результатів. Додатково на результати можуть впливати вибір гіперпараметрів моделі, порогів прийняття рішення, спосіб балансування даних і конкретна конфігурація навчального процесу. Отже, результати експериментів слід розглядати як підтвердження перспективності запропонованих методів, але не як вичерпну характеристику їх можливої ефективності в усіх умовах реального використання.

Необхідно також враховувати обмеження зовнішньої валідності. Ефективність методів перевірено на англomовному користувацькому контенті платформ Reddit і Twitter; застосування до інших мов, платформ або клінічних вибірок потребує донавчання моделей на відповідних даних, що не змінює самих методів, а лише їх параметрів. Особливо це важливо для задачі виявлення ПТСР, де мовні прояви травматичного досвіду можуть залежати від культурного контексту, типу пережитої травми, часу після травматичної події, індивідуальної здатності до вербалізації переживань і соціально прийнятних форм самовираження.

Важливим обмеженням усіх трьох методів є те, що вони не призначені для автономного використання як інструменти клінічної діагностики. Запропоновані моделі та правила аналізу орієнтовані на інтелектуальний скринінг, попередню аналітичну оцінку та виявлення ризиків на основі цифрового текстового сліду користувача. Їх застосування може бути корисним у системах психологічного моніторингу, підтримки фахівців, первинного сортування звернень або аналітичного супроводу соціальних служб, однак остаточні висновки щодо психічного стану людини, наявності ПТСР чи рівня ризику соціально небезпечної поведінки повинні прийматися лише за участю кваліфікованого фахівця з урахуванням клінічного, поведінкового та соціального контексту.

Обмеження, пов'язані зі складом набору даних першого методу. Позитивний клас сформовано з двох джерел (584 записи сабрeдиту ptsd набору «Human Stress Prediction» і 536 записів набору «Aya PTSD»), а

негативний – лише з «Human Stress Prediction». Тому близько половини позитивних прикладів відрізняються від негативних ще й джерелом, і модель могла частково навчитися розрізняти джерело даних (довжину, стиль, форматування тексту), а не лише ознаки ПТСР; окремого контролю цього ефекту (навчання на одному джерелі, метрики окремо за джерелами) не виконували. Мітка «ПТСР» у наборі «Human Stress Prediction» визначається тематикою сабредиту. Нецільовий клас утворюють дописи інших спільнот, лише частина яких (anxiety, stress) стосується психічних розладів, а дописи спільнот постраждалих від насильства можуть містити наративи травматичного досвіду. Частоту хибних спрацьовувань окремо за сабредитами не оцінювали, тому відмежування ПТСР від інших розладів підтверджено лише для нецільового класу в цілому.

Обмеження протоколу оцінювання першого методу. Метрики отримано на валідаційній вибірці (20 % збалансованого набору, 440 записів), яку використовували і для контролю навчання; окремої тестової вибірки не виділяли, а результат одержано за один запуск навчання без оцінки розкиду між запусками. Набір «Human Stress Prediction» складається з фрагментів дописів (поля `post_id`, `sentence_range`), а поділ виконано випадково за записами, тому фрагменти одного допису могли потрапити і до навчальної, і до валідаційної вибірки. Отже, наведені значення можуть бути дещо завищеними. ROC-криву рис. 4.3 побудовано за бінарними рішеннями, тому рангову ROC-AUC моделі DeBERTa в роботі не визначено. Порівняння з базовими моделями (BERT, RoBERTa) на тому самому розбитті та з варіантом нецільового класу лише з текстів без ознак розладів не проводили, тому внесок кожної зі складових методу окремо не оцінено.

Обмеження, пов'язані з наборами даних другого методу. Цільові тексти всіх п'яти розладів узято з набору «COMSYS-T1» (твіти), а частину нецільового класу («здорові» тексти) – з набору «Depression: Reddit Dataset (Cleaned)». Відмінності платформ за довжиною і стилем текстів є додатковим сигналом, який може частково пояснювати дуже високі значення метрик

окремих моделей (наприклад, для «Narcissistic Disorder» і «Panic Disorder»). Процедуру розмітки набору «COMSYS-T1» його упорядники детально не описали. У наборах даних другого методу немає класу ПТСР, тому зв'язок виявлюваних розладів саме з ПТСР (їх «супутність») у роботі не перевірявся.

Обмеження протоколу оцінювання другого методу. Тестові вибірки бінарних моделей містять від 45 до 83 записів, тому довірчі інтервали метрик широкі (для Accuracy 0,918 на 61 записі 95 % інтервал Вілсона становить приблизно 0,82–0,96), а різниці на рівні сотих між конфігураціями та з опублікованими аналогами не є статистично значущими. Багатокласову модель оцінено на спільній вибірці зі 157 текстів, а бінарні – кожен на власній збалансованій вибірці, тому твердження про відсутність втрати точності порівняно з багатокласовою класифікацією має орієнтовний характер. Варіанти з адаптивним порогом і без нього відрізняються також розміром пачки та кількістю епох, а пороги за індексом Юдена визначено за оцінками на навчальній вибірці (алгоритм 3.5); тому приріст Macro F1-score з 0,925 до 0,9296 не можна приписати лише порогованню.

Обмеження третього методу. Параметри індексу ризику (ширина вікна $W = 30$ днів, пороги 0,30 і 0,50, прирости 0,2 і 0,3) задано з наведених у п. 2.4 міркувань; аналіз чутливості індексу до цих параметрів не виконували. За малого наповнення вікна індикатори (2.17) обчислюються за одним-двома повідомленнями, тому поодинокі, зокрема хибнопозитивні, рішення може різко змінити індекс; для практичного застосування доцільно задавати мінімальну кількість повідомлень у вікні. Ознаки табл. 2.1 до індексу не входять і експериментально не оцінювались. Набір даних п. 3.5.3 містить лише мітку депресивного чи недепресивного твіту і не містить ні ознак ПТСР, ні даних про поведінку користувачів, тому перевірено працездатність механізму агрегації, а не прогностичну здатність індексу.

Пояснюваність рішень першого методу продемонстровано на окремих прикладах; кількісної оцінки правдивості та стабільності пояснень не виконували. Модель із квантовим шаром досліджено в одному запуску для

кожної конфігурації, а Dropout із ймовірністю 0,6–0,75 після шару з двома виходами є недоліком цієї конфігурації; тому її результати мають попередній характер.

Таким чином, запропоновані методи демонструють перспективність для автоматизованого аналізу психічного стану громадян за користувацьким контентом, однак мають низку обмежень, пов'язаних із природою використаних даних, складністю предметної області, неповнотою розмітки, клінічною близькістю окремих станів, варіативністю мовних проявів та межами практичного застосування результатів. Урахування цих обмежень не знижує наукової цінності проведеного дослідження, а навпаки, дає змогу коректно визначити межі застосовності розроблених підходів і окреслити подальші напрями їх удосконалення, зокрема шляхом розширення та уніфікації наборів даних, уточнення системи ознак, посилення інтерпретованості моделей і проведення валідації на нових вибірках користувацького контенту.

4.7. Висновки до розділу 4

У розділі експериментально досліджено запропонований підхід до вирішення задачі виявлення проявів ПТСР в користувацькому текстовому контенті. Підхід відрізняється від існуючих збільшеним фокусуванням на специфічних для ПТСР контекстних залежностях, й покращеній роздільності від проявів інших психічних розладів.

Покращення роздільності від проявів інших психічних розладів досягнуто за рахунок включення навчальних текстів з проявами інших психічних розладів до нецільового класу набору даних. Використання при навчанні набору даних, в якому у протилежній категорії знаходились дані не лише психологічно здорових людей, а і дані людей з іншими психічними розладами, може сприяти мінімізації плутанини з іншими психічними розладами. Це було досягнуто шляхом створення власного навчального

набору даних із 3374 записів, з яких для навчання й валідації використано збалансовану вибірку з 2200 записів.

Розроблений метод для аналізу прояву ПТСР в текстовому контенті перетворює вхідний текст за допомогою навченої контекстно-орієнтованої нейромережевої моделі з токенизатором у модельну оцінку прояву ПТСР в користувацькому контенті. Запропонований підхід дозволяє не лише визначати прояв ПТСР в текстовому контенті, а й візуально інтерпретувати отримані результати.

Було досліджено ефективність розробленого методу шляхом розробки відповідного програмного забезпечення. При навчанні нейромережі описаної архітектури, були отримані результати за метриками: Accuracy 0,934, Precision 0,948, F1 0,933 на валідаційній вибірці (440 записів). Опубліковані для схожих задач значення (збалансована точність 0,804 [17], F1 0,810 [19], AUC 0,690 [18] і 0,740 [34]) отримано на інших наборах даних і наведено лише як контекст.

Завдяки інтеграції найсучасніших методів глибокого навчання з передовою обробкою природної мови, розроблений підхід дозволяє ідентифікувати мовні та поведінкові маркери, пов'язані з такими станами, як депресія, тривога та нарцисичний розлад.

Удосконалений метод виявлення психічних розладів за аналізом користувацького контенту дозволяє отримати в моделях покращену здатність розрізняти специфічні ознаки для кожного виду психічного розладу, за рахунок навчання на модифікованих множинах текстових даних, що зменшує ймовірність плутанини між розладами, оскільки модель вчиться розрізняти їхні характерні ознаки. Одержано високі результати точності та повноти виявлення психічних розладів розробленим методом: при багатокласовій класифікації значення F1 макрометрики одержано 86,2 %, при підході на основі множини бінарних класифікаторів значення F1 макрометрики одержано 92,5 % (92,96 % з адаптивним порогуванням), тоді як опубліковані на інших корпусах аналоги мають значення F1 макрометрики 64,5 % [42] і 89,65 % [43].

Одержані результати не тільки демонструють ефективність моделей нейронних мереж у виявленні тонких лінгвістичних сигналів, пов'язаних з проблемами психічного здоров'я, але й дають зрозуміти ключові особливості, які можуть служити ранніми індикаторами погіршення психічного здоров'я. Ця робота сприяє зростанню сфери обчислювального психічного здоров'я, пропонуючи масштабований, неінвазивний і економічно ефективний інструмент для моніторингу психічного благополуччя в режимі реального часу, адже зростаюча поширеність розладів психічного здоров'я в усьому світі вимагає інноваційних підходів для раннього виявлення та втручання, що визначено критеріями Програми розвитку ООН у цілі сталого розвитку №3 «Міцне здоров'я».

Для методу розширення множини ознак реалізовано часовий індекс ризику, який інтегрує рішення щодо ознак ПТСР, інших психічних розладів та негативного емоційного контексту. На профілях користувачів набору даних із часовими мітками показано, що індекс розрізняє профілі з епізодичними сплесками, без коливань і зі стійко високим рівнем (рис. 4.17–4.19), а його складові пояснювані за (2.18). Виявлено джерела хибнопозитивних рішень (сарказм, ігровий контекст) і чутливість індексу до малого наповнення вікна; прогностичну точність індексу щодо реальної поведінки не оцінювали.

Окрема модель із квантовим шаром (Accuracy 0,85, AUC 0,88) поступається моделі DeBERTa і розглядається як пошуковий експеримент; різницю в 1 відсотковий пункт порівняно з моделлю без квантового шару отримано за один запуск для різних значень Dropout, і вона не є підтвердженням ефекту квантового шару.

Обмеженнями запропонованого підходу є певні труднощі з класифікацією психічних розладів «Anxiety Disorder» та «Depression», що пов'язано з їх клінічною подібністю. Повний перелік обмежень, зокрема щодо складу наборів даних і протоколу оцінювання, наведено в п. 4.6.

Результати експериментальних досліджень, наведені в розділі, опубліковано в працях [158–166], а розроблені програмні засоби захищено свідоцтвами про реєстрацію авторського права [167, 168].

ВИСНОВКИ

У результаті виконання дисертаційного дослідження було розв'язано актуальну науково-прикладну задачу аналізу психічного стану громадян із посттравматичним стресовим розладом за інтелектуальним аналізом користувацького контенту. Розроблені методи дають змогу ефективно виявляти ознаки ПТСР у текстовому контенті, ідентифікувати супутні психічні розлади та формувати скринінгову оцінку ризику соціально небезпечної поведінки на основі аналізу мовних, емоційно-семантичних і поведінкових патернів цифрової активності користувачів. Це створює підґрунтя для використання запропонованого підходу в системах психологічного моніторингу, соціальної підтримки та раннього виявлення осіб, які потребують допомоги.

Розроблені методи дозволили досягти мети дисертаційного дослідження: підвищити точність виявлення ознак посттравматичного стресового розладу і супутніх психічних розладів та розширити можливості оцінювання ризиків соціально небезпечної поведінки громадян за інтелектуальним аналізом користувацького контенту засобами контекстно-орієнтованих нейромережових моделей і розширення простору ознак. Кількісні результати експериментів наведено у розділі 4; нижче подано результати за кожним із поставлених завдань.

1. Проаналізовано сучасні методи, моделі та засоби автоматизованого виявлення ПТСР та інших психічних розладів за текстовим користувацьким контентом. Встановлено, що відомі рішення з виявлення ПТСР побудовано переважно на клінічних інтерв'ю та наративах вузьких груп і вони не передбачають контрольованого відмежування ознак ПТСР від тематично суміжних станів на рівні дописів; моделі виявлення психічних розладів орієнтовано на один домінуючий клас; поведінкові ризики оцінюються за окремими ознаками без урахування їх часової динаміки.

2. Удосконалено метод аналізу прояву посттравматичного стресового розладу в користувацькому контенті, який відрізняється від існуючих формуванням спеціалізованого навчального набору даних, нецільовий клас якого містить тексти тематично суміжних спільнот, зокрема з проявами інших психічних розладів, у поєднанні з контекстно-позиційним представленням тексту, що дало змогу підвищити точність виявлення ознак ПТСР. На валідаційній вибірці (440 записів) отримано Accurasy 0,934, Precision 0,948, F1-score 0,933 і специфічність 0,950. Пояснюваність рішень продемонстровано засобами Transformers Interpret і BertViz [143, 144].

3. Удосконалено метод виявлення психічних розладів за аналізом користувацького контенту, який відрізняється від існуючих декомпозицією задачі на множину незалежних бінарних класифікаторів з окремо сформованими навчальними вибірками та індивідуальними порогами прийняття рішення, що не потребує вибору одного домінуючого розладу і не знижує точність порівняно з багатокласовою класифікацією. Конфігурація з індивідуальними порогами за індексом Юдена забезпечила Macro Accurasy 0,9316, Macro Precision 0,9338, Macro Recall 0,9296 та Macro F1-score 0,9296 проти Macro F1-score 0,862 багатокласової моделі; Accurasy окремих бінарних моделей становить від 0,87 до 1,00 (табл. 4.8). Одночасне виявлення кількох розладів забезпечено архітектурно, але потребує перевірки на багатомітково розмічених даних.

4. Вперше розроблено метод розширення множини ознак користувацького контенту для попередження соціально небезпечної поведінки громадян із посттравматичним стресовим розладом, який відрізняється від існуючих підходів інтеграцією оцінок прояву ПТСР, інших психічних розладів та емоційного контексту повідомлень у пояснюваний часовий індекс ризику на рівні користувача, що дає змогу формувати скринінгову оцінку ризику соціально небезпечної поведінки. Працездатність механізму інтеграції та часової агрегації показано на профілях користувачів

набору даних із часовими мітками (рис. 4.17–4.19); прогностичну точність індексу щодо реальної поведінки не оцінювали.

5. Розроблено програмну систему інтелектуального аналізу користувацького контенту, компоненти якої захищено свідоцтвами про реєстрацію авторського права [167, 168], а програмний код розміщено у відкритому репозиторії (Додаток В), і виконано експериментальне дослідження запропонованих методів. Додатково досліджено окрему модель із квантовим шаром (Assurasy 0,85), яка до складу методів не входить і розглядається як пошуковий експеримент.

Межі застосування результатів. Методи перевірено на англomовному користувацькому контенті платформ Reddit і Twitter; мітки наборів даних відображають тематику онлайн-спільнот, а не клінічні діагнози. Методи призначені для скринінгу мовних ознак у текстах, не є засобом клінічної діагностики чи прогнозування поведінки конкретної особи і потребують участі фахівця в ухваленні рішень. Обмеження наборів даних і протоколу оцінювання наведено в п. 4.6.

Практичне значення отриманих результатів полягає у програмній реалізації комплексного підходу до інтелектуального аналізу користувацького контенту з метою виявлення ознак посттравматичного стресового розладу та скринінгового оцінювання ризику соціально небезпечної поведінки. Реалізована система дозволяє автоматизовано обробляти великі обсяги текстової інформації, формувати модельну оцінку прояву ознак ПТСР у текстах, виявляти супутні психічні стани, аналізувати типові патерни мовлення та поведінки з числовими характеристиками їх виразності, а також візуалізувати результати роботи моделей. Це сприяє підвищенню прозорості, інтерпретованості та довіри до отриманих результатів з боку фахівців і кінцевих користувачів.

Результати дисертаційної роботи впроваджено в освітній процес Хмельницького національного університету при викладанні дисциплін

«Методи та системи штучного інтелекту» й «Моделі та методи текстової аналітики», а також у діяльність ГО «ІТ Кластер м. Хмельницького», ПП «Авіві» та ТОВ «Стфалкон» (Додаток Б).

Напрями подальших досліджень: оцінювання першого методу на незалежній тестовій вибірці з груповим розбиттям і порівняння з базовими моделями; формування багатомітково розміченої вибірки для перевірки одночасного виявлення кількох розладів; донавчання моделей на україномовних даних; включення ознак табл. 2.1 до індексу ризику, аналіз його чутливості до параметрів і зовнішня валідація за участі фахівців; врахування сарказму та ігрового контексту.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Prykhodko I. I. Program of psychological rehabilitation of the National Guard of Ukraine military personnel participated in combat actions. *International Journal of Science Annals*. 2018. Vol. 1, no. 1-2. P. 34–42. URL: <https://doi.org/10.26697/ijsa.2018.1-2.05> (дата звернення: 31.03.2026).
2. Psychological well-being of Ukrainian students three months after the emerge of full-scale war / I. Pinchuk et al. *Psychiatria Polska*. 2024. Vol. 58, no. 1. P. 121–151. URL: <https://doi.org/10.12740/pp/177073> (дата звернення: 31.03.2026).
3. Prevalence of stress, anxiety, and symptoms of post-traumatic stress disorder among Ukrainians after the first year of Russian invasion: a nationwide cross-sectional study / O. Lushchak et al. *The Lancet Regional Health - Europe*. 2024. Vol. 36. P. 100773. URL: <https://doi.org/10.1016/j.lanepe.2023.100773> (дата звернення: 31.03.2026).
4. Mental health in Ukraine in 2023 / D. Martsenkovskyi et al. *European Psychiatry*. 2024. Vol. 67, no. 1. URL: <https://doi.org/10.1192/j.eurpsy.2024.12> (дата звернення: 31.03.2026).
5. PTSD in military personnel: diagnosis, treatment and support / I. Klymenko et al. *Revista Amazonia Investiga*. 2024. Vol. 13, no. 74. P. 286–298. URL: <https://doi.org/10.34069/ai/2024.74.02.24> (дата звернення: 31.03.2026).
6. Post-traumatic stress disorder: clinical and laboratory changes and potential for immune disorders / S. Zubchenko et al. *Proceeding of the Shevchenko Scientific Society. Medical Sciences*. 2023. Vol. 71, no. 1. URL: <https://doi.org/10.25040/ntsh2023.01.11> (дата звернення: 31.03.2026).
7. Anisimov A. V., Marchenko O. O., Nasirov E. M., Taranukha V. Y. Tuning BERT-based Models for Sentiment Analysis of Texts. *Cybernetics and Systems Analysis*. 2026. Vol. 62, No. 2. P. 196–203. URL: <https://doi.org/10.1007/s10559-026-00857-4> (дата звернення: 25.09.2026).

8. Marchenko O. O., Nasirov E. M., Volosheniuk D. O. Building of the Ukrainian-language Training Dataset for Determining the Sentiment Analysis of Texts. *Control Systems and Computers*. 2024. No. 4 (308). P. 34–38. URL: <https://doi.org/10.15407/csc.2024.04.034> (дата звернення: 25.09.2026).
9. Gramyak R., Lipyanina-Goncharenko H., Sachenko A., Lendyuk T., Zahorodnia D. Intelligent Method of a Competitive Product Choosing based on the Emotional Feedbacks Coloring. *CEUR Workshop Proceedings*. 2021. Vol. 2853. P. 346–357. URL: <https://ceur-ws.org/Vol-2853/paper31.pdf> (дата звернення: 25.09.2026).
10. Федоричко А. С., Висоцька В. А., Линник Р. О. Інформаційна технологія класифікації текстів для виявлення ознак ПТСР на основі технологій глибинного навчання та NLP. *Біоніка інтелекту*. 2026. № 1 (104). С. 45–55. URL: [https://doi.org/10.30837/bi.2026.1\(104\).06](https://doi.org/10.30837/bi.2026.1(104).06) (дата звернення: 26.09.2026).
11. Корбило Р. Я., Грицюк Ю. І. Система оцінювання психологічного стану військовослужбовців методами оброблення природної мови. *Науковий вісник НЛТУ України*. 2025. Т. 35, № 6. С. 137–153. URL: <https://doi.org/10.36930/40350617> (дата звернення: 25.09.2026).
12. Barmak O., Manziuk E., Krak I. Using piecewise hyper linear classification in multidimensional feature space for text content. *2019 IEEE 14th International Conference on Computer Sciences and Information Technologies (CSIT)*, Lviv, Ukraine, 17–20 September 2019. 2019. P. 119–123. URL: <https://doi.org/10.1109/stc-csit.2019.8929798> (дата звернення: 26.09.2026).
13. Chaban O., Manziuk E. Enhancing medical NLI with integrated domain knowledge and sentiment analysis. *CEUR Workshop Proceedings*. 2024. Vol. 3790. P. 262–272. URL: <https://ceur-ws.org/Vol-3790/paper23.pdf> (дата звернення: 26.09.2026).
14. Mazurets O., Barmak O., Krak I., Manziuk E., Bahrii R. Method for Adaptive Semantic Testing of Educational Materials Level of Knowledge. *Lecture Notes in Computational Intelligence and Decision Making. Lecture Notes on Data*

Engineering and Communications Technologies. Cham : Springer, 2021. P. 491–506. URL: https://doi.org/10.1007/978-3-030-82014-5_33 (дата звернення: 26.09.2026).

15. Radiuk P., Barmak O., Manziuk E., Krak I. Explainable Deep Learning: A Visual Analytics Approach with Transition Matrices. *Mathematics*. 2024. Vol. 12, no. 7. Art. 1024. URL: <https://doi.org/10.3390/math12071024> (дата звернення: 26.09.2026).

16. Radiuk P., Barmak O., Manziuk E., Krak I. Explainable Artificial Intelligence: Transitioning DL Model Decisions to User-Understandable Features in Healthcare. *CEUR Workshop Proceedings*. 2024. Vol. 3777. P. 185–199. URL: <https://ceur-ws.org/Vol-3777/paper12.pdf> (дата звернення: 26.09.2026).

17. Detecting Presence of PTSD Using Sentiment Analysis From Text Data / J. Sawalha et al. *Frontiers in Psychiatry*. 2022. Vol. 12. URL: <https://doi.org/10.3389/fpsy.2021.811392> (дата звернення: 31.03.2026).

18. Interdisciplinary approach to identify language markers for post-traumatic stress disorder using machine learning and deep learning / R. Quillivic et al. *Scientific Reports*. 2024. Vol. 14, no. 1. URL: <https://doi.org/10.1038/s41598-024-61557-7> (дата звернення: 31.03.2026).

19. AI and narrative embeddings detect PTSD following childbirth via birth stories / A. Bartal et al. *Scientific Reports*. 2024. Vol. 14, no. 1. URL: <https://doi.org/10.1038/s41598-024-54242-2> (дата звернення: 31.03.2026).

20. Artificial Intelligence for Analyzing Mental Health Disorders in Social Media: A Quarter-Century Narrative Review of Progress and Challenges (Preprint) / D. Owen et al. *Journal of Medical Internet Research*. 2024. URL: <https://doi.org/10.2196/59225> (дата звернення: 31.03.2026).

21. Natural language processing for mental health interventions: a systematic review and research framework / M. Malgaroli et al. *Translational Psychiatry*. 2023. Vol. 13, no. 1. URL: <https://doi.org/10.1038/s41398-023-02592-2> (дата звернення: 31.03.2026).

22. Characterisation of mental health conditions in social media using Informed Deep Learning / G. Gkotsis et al. *Scientific Reports*. 2017. Vol. 7, no. 1. URL: <https://doi.org/10.1038/srep45141> (дата звернення: 31.03.2026).

23. A deep learning model for detecting mental illness from user content on social media / J. Kim et al. *Scientific Reports*. 2020. Vol. 10, no. 1. URL: <https://doi.org/10.1038/s41598-020-68764-y> (дата звернення: 31.03.2026).

24. Machine learning models predict PTSD severity and functional impairment: A personalized medicine approach for uncovering complex associations among heterogeneous symptom profiles. / A. H. Park et al. *Psychological Trauma: Theory, Research, Practice, and Policy*. 2025. URL: <https://doi.org/10.1037/tra0001602> (дата звернення: 31.03.2026).

25. Malgaroli M., Schultebrucks K. Artificial Intelligence and Posttraumatic Stress Disorder (PTSD). *European Psychologist*. 2020. Vol. 25, no. 4. P. 272–282. URL: <https://doi.org/10.1027/1016-9040/a000423> (дата звернення: 31.03.2026).

26. Machine Learning Driven Mental Stress Detection on Reddit Posts Using Natural Language Processing / S. Inamdar et al. *Human-Centric Intelligent Systems*. 2023. URL: <https://doi.org/10.1007/s44230-023-00020-8> (дата звернення: 31.03.2026).

27. Arowosegbe A., Oyelade T. Application of Natural Language Processing (NLP) in Detecting and Preventing Suicide Ideation: A Systematic Review. *International Journal of Environmental Research and Public Health*. 2023. Vol. 20, no. 2. P. 1514. URL: <https://doi.org/10.3390/ijerph20021514> (дата звернення: 31.03.2026).

28. A survey on detecting mental disorders with natural language processing: Literature review, trends and challenges / A. Montejo-Ráez et al. *Computer Science Review*. 2024. Vol. 53. P. 100654. URL: <https://doi.org/10.1016/j.cosrev.2024.100654> (дата звернення: 31.03.2026)..

29. Global prevalence and associated risk factors of posttraumatic stress disorder during COVID-19 pandemic: A meta-analysis / N. Yunitri et al. *International Journal of Nursing Studies*. 2022. Vol. 126. P. 104136. URL: <https://doi.org/10.1016/j.ijnurstu.2021.104136> (дата звернення: 26.09.2026).
30. AVEC'19 / F. Ringeval et al. *MM '19: The 27th ACM International Conference on Multimedia*, Nice France. New York, NY, USA, 2019. URL: <https://doi.org/10.1145/3343031.3350550> (дата звернення: 19.09.2026).
31. A machine learning approach to modeling PTSD and difficulties in emotion regulation / N. M. Christ et al. *Psychiatry Research*. 2021. Vol. 297. P. 113712. URL: <https://doi.org/10.1016/j.psychres.2021.113712> (дата звернення: 19.09.2026).
32. Identifying Women with Post-Delivery Posttraumatic Stress Disorder using Natural Language Processing of Personal Childbirth Narratives / A. BARTAL et al. *American Journal of Obstetrics & Gynecology MFM*. 2023. Vol. 5, no. 3. Art. 100834. URL: <https://doi.org/10.1016/j.ajogmf.2022.100834> (дата звернення: 19.09.2026).
33. DSM-5-TR: Rationale, Process, and Overview of Changes / M. B. First et al. *Psychiatric Services*. 2023. URL: <https://doi.org/10.1176/appi.ps.20220334> (дата звернення: 19.05.2025).
34. Development and Validation of a Machine Learning Prediction Model of Posttraumatic Stress Disorder After Military Deployment / S. Papini et al. *JAMA Network Open*. 2023. Vol. 6, no. 6. P. e2321273. URL: <https://doi.org/10.1001/jamanetworkopen.2023.21273> (дата звернення: 19.09.2026).
35. Identifying posttraumatic stress disorder staging from clinical and sociodemographic features: a proof-of-concept study using a machine learning approach / L. F. Ramos-Lima et al. *Psychiatry Research*. 2022. Vol. 311. P. 114489. URL: <https://doi.org/10.1016/j.psychres.2022.114489> (дата звернення: 19.09.2026).

36. Uban A.-S., Chulvi B., Rosso P. An emotion and cognitive based analysis of mental health disorders from social media data. *Future Generation Computer Systems*. 2021. Vol. 124. P. 480–494. URL: <https://doi.org/10.1016/j.future.2021.05.032> (дата звернення: 19.09.2026).
37. Understanding mental health issues in different subdomains in social networking services: computational analysis of text-based reddit posts (Preprint) / S. Kim et al. *Journal of Medical Internet Research*. 2023. URL: <https://doi.org/10.2196/49074> (дата звернення: 19.09.2026).
38. Detecting mental distresses using social behavior analysis in the context of COVID-19: A survey / S. Dhelim et al. *ACM Computing Surveys*. 2023. Vol. 55, no. 14s. Art. 306. P. 1–30. URL: <https://doi.org/10.1145/3589784> (дата звернення: 26.09.2026).
39. A lexicon-based approach to detecting suicide-related messages on Twitter / S. M. Sarsam et al. *Biomedical Signal Processing and Control*. 2021. Vol. 65. P. 102355. URL: <https://doi.org/10.1016/j.bspc.2020.102355> (дата звернення: 19.09.2026).
40. Hierarchical Multiscale Recurrent Neural Networks for Detecting Suicide Notes / A. M. Schoene et al. *IEEE Transactions on Affective Computing*. 2023. Vol. 14, no. 1. P. 153–164. URL: <https://doi.org/10.1109/taffc.2021.3057105> (дата звернення: 19.09.2026).
41. Detection of Suicide Ideation in Social Media Forums Using Deep Learning / M. M. Tadesse et al. *Algorithms*. 2019. Vol. 13, no. 1. P. 7. URL: <https://doi.org/10.3390/a13010007> (дата звернення: 19.09.2026).
42. Detection of Mental Health from Reddit via Deep Contextualized Representations / Z. Jiang et al. *Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis*, Online. Stroudsburg, PA, USA, 2020. URL: <https://doi.org/10.18653/v1/2020.louhi-1.16> (дата звернення: 19.09.2026).
43. Mental Illness Classification on Social Media Texts Using Deep Learning and Transfer Learning / M. Arif et al. *Computación y Sistemas*. 2024.

Vol. 28, no. 2. URL: <https://doi.org/10.13053/cys-28-2-4873> (дата звернення: 19.05.2025).

44. Deep learning and machine learning in psychiatry: a survey of current progress in depression detection, diagnosis and treatment / M. Squires et al. *Brain Informatics*. 2023. Vol. 10, no. 1. Art. 10. URL: <https://doi.org/10.1186/s40708-023-00188-6> (дата звернення: 26.09.2026).

45. Dynamic changes and future trend predictions of the global burden of anxiety disorders: analysis of 204 countries and regions from 1990 to 2021 and the impact of the COVID-19 pandemic / S. Chen et al. *eClinicalMedicine*. 2025. Vol. 79. P. 103014. URL: <https://doi.org/10.1016/j.eclinm.2024.103014> (дата звернення: 19.09.2026).

46. Global Scientific Trends on Anxiety Disorders in the Early Twenty-First Century: A Data-Driven Bibliometric Analysis / Z. Wang et al. *Depression and Anxiety*. 2025. Vol. 2025, no. 1. URL: <https://doi.org/10.1155/da/5541514> (дата звернення: 19.05.2025).

47. Predicting Diagnostic Progression to Schizophrenia or Bipolar Disorder via Machine Learning / L. Hansen et al. *JAMA Psychiatry*. 2025. URL: <https://doi.org/10.1001/jamapsychiatry.2024.4702> (дата звернення: 19.05.2025).

48. Hirakawa H., Hori H., Terao T. Editorial: Bipolar disorder and cognition: cognitive decline and dementia. *Frontiers in Psychiatry*. 2025. Vol. 16. URL: <https://doi.org/10.3389/fpsy.2025.1567775> (дата звернення: 19.05.2025).

49. Malhi G. S., Szymaniak K., Bell E. Does ADHD treatment inefficacy question its diagnostic validity?. *The Lancet Psychiatry*. 2025. Vol. 12, no. 5. P. 329. URL: [https://doi.org/10.1016/s2215-0366\(25\)00045-8](https://doi.org/10.1016/s2215-0366(25)00045-8) (дата звернення: 19.05.2025).

50. Artificial intelligence in ADHD: a global perspective on research hotspots, trends and clinical applications / X. Wang et al. *Frontiers in Human Neuroscience*. 2025. Vol. 19. URL: <https://doi.org/10.3389/fnhum.2025.1577585> (дата звернення: 19.05.2025).

51. A Novel Co-Training-Based Approach for the Classification of Mental Illnesses Using Social Media Posts / S. Tariq et al. *IEEE Access*. 2019. Vol. 7. P. 166165–166172. URL: <https://doi.org/10.1109/access.2019.2953087> (дата звернення: 26.09.2026).

52. Madububambachu U., Ukpebor A., Ihezue U. Machine Learning Techniques to Predict Mental Health Diagnoses: A Systematic Literature Review. *Clinical Practice & Epidemiology in Mental Health*. 2024. Vol. 20, no. 1. URL: <https://doi.org/10.2174/0117450179315688240607052117> (дата звернення: 19.09.2026).

53. Pourkeyvan A., Safa R., Sorourkhah A. Harnessing the Power of Hugging Face Transformers for Predicting Mental Health Disorders in Social Networks. *IEEE Access*. 2024. Vol. 12. P. 28025–28035 URL: <https://doi.org/10.1109/access.2024.3366653> (дата звернення: 19.09.2026).

54. Sutranggono A. N., Riyanarto Sarno, Imam Ghozali. Multi-Class Multi-Level Classification of Mental Health Disorders Based on Textual Data from Social Media. *Journal of Information and Communication Technology*. 2024. Vol. 23, no. 1. P. 77–104. URL: <https://doi.org/10.32890/jict2024.23.1.4> (дата звернення: 19.05.2025).

55. A Hybrid Transformer Architecture for Multiclass Mental Illness Prediction using Social Media Text / A. Karamat et al. *IEEE Access*. 2025. Vol. 13. P. 12148–12167. URL: <https://doi.org/10.1109/access.2024.3519308> (дата звернення: 19.09.2026).

56. Carreras Timoneda J., Vallejo Vera S. BERT, RoBERTa or DeBERTa? Comparing Performance Across Transformer Models in Political Science Text. *The Journal of Politics*. 2024. URL: <https://doi.org/10.1086/730737> (дата звернення: 19.05.2025).

57. Liu Y., Ott M., Goyal N., Du J., Joshi M., Chen D., Levy O., Lewis M., Zettlemoyer L., Stoyanov V. RoBERTa: A Robustly Optimized BERT Pretraining Approach. arXiv preprint arXiv:1907.11692. 2019. URL: <https://arxiv.org/abs/1907.11692> (дата звернення: 01.10.2026).

58. Nabiilah G. Z. Effectiveness Analysis of RoBERTa and DistilBERT in Emotion Classification Task on Social Media Text Data. *Engineering, Mathematics and Computer Science Journal (EMACS)*. 2025. Vol. 7, no. 1. P. 45–

50. URL: <https://doi.org/10.21512/emacsjournal.v7i1.12618> (дата звернення: 31.03.2026).

59. Ji S., Zhang T., Ansari L., Fu J., Tiwari P., Cambria E. MentalBERT: Publicly Available Pretrained Language Models for Mental Healthcare. Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC 2022). Marseille, France, 2022. P. 7184–7190. URL: <https://aclanthology.org/2022.lrec-1.778> (дата звернення: 01.10.2026).

60. MelBERT: Metaphor Detection via Contextualized Late Interaction using Metaphorical Identification Theories / M. Choi et al. *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Online*. Stroudsburg, PA, USA, 2021. URL: <https://doi.org/10.18653/v1/2021.naacl-main.141> (дата звернення: 31.03.2026).

61. Devlin J., Chang M.-W., Lee K., Toutanova K. 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 (NAACL-HLT 2019)*. Minneapolis, MN, USA, 2019. Vol. 1. P. 4171–4186. URL: <https://doi.org/10.18653/v1/N19-1423> (дата звернення: 26.09.2026).

62. niss.gov.ua. Підтримка ментального здоров'я в часи війни. URL: <https://niss.gov.ua/news/komentari-ekspertiv/pidtrymka-mentalnoho-zdorovya-v-chasy-viyny> (дата звернення: 19.05.2025).

63. Формування спільноти фахівців для більш ефективної допомоги людям із ПТСР. URL: <https://www.kmu.gov.ua/news/formuvannia-spilnoty-fakhivtsiv-dlia-bilsh-efektyvnoi-dopomohy-liudiam-iz-ptsr-vidbulas-ustanovcha-konferentsiia> (дата звернення: 19.05.2025).

64. Система у сфері психічного здоров'я та психосоціальної підтримки в Україні. URL: [https://howareu.com/static-objects/howareu/media/Posibnuki/2024-06-](https://howareu.com/static-objects/howareu/media/Posibnuki/2024-06-06%20%D0%A6%D0%9C%20%D0%B2%D0%B5%D1%80%D1%81%D1%96%D1%8F%202.0.pdf)

[06%20%D0%A6%D0%9C%20%D0%B2%D0%B5%D1%80%D1%81%D1%96%D1%8F%202.0.pdf](https://howareu.com/static-objects/howareu/media/Posibnuki/2024-06-06%20%D0%A6%D0%9C%20%D0%B2%D0%B5%D1%80%D1%81%D1%96%D1%8F%202.0.pdf) (дата звернення: 19.05.2025).

65. Zbitnieva O., Chaban O., Omelyanovich V. "A target model of the mental health and psychosocial support system" As a response to the devastating impact of war in Ukraine. *Psychosomatic Medicine and General Practice*. 2024. Vol. 9, no. 3. URL: <https://doi.org/10.26766/pmgrp.v9i3.539> (дата звернення: 19.09.2026).

66. Уніфікований клінічний протокол первинної та спеціалізованої медичної допомоги «Гостра реакція на стрес. Посттравматичний стресовий розлад. Порушення адаптації»: затв. наказом МОЗ України від 19.07.2024 № 1265. URL: https://moz.gov.ua/storage/uploads/ec4ae01d-d0d3-4c0a-bf92-3cefbe633be/dn_1265_19072024_dod.pdf (дата звернення: 26.09.2026).

67. National Institute for Health and Care Excellence. Post-traumatic stress disorder : NICE guideline NG116. London : NICE, 2018. URL: <https://www.nice.org.uk/guidance/ng116> (дата звернення: 26.09.2026).

68. Brooks S. K., Greenberg N. Recurrence of post-traumatic stress disorder: systematic review of definitions, prevalence and predictors. *BMC Psychiatry*. 2024. Vol. 24, no. 1. URL: <https://doi.org/10.1186/s12888-023-05460-x> (дата звернення: 19.05.2025).

69. Prevalence of post-traumatic stress disorder and depressive symptoms among civilians residing in armed conflict-affected regions: a systematic review and meta-analysis / S. H. Ahmed et al. *General Psychiatry*. 2024. Vol. 37, no. 3. P. e101438. URL: <https://doi.org/10.1136/gpsych-2023-101438> (дата звернення: 19.05.2025).

70. Bisson J. I., Cosgrove S., Lewis C., Roberts N. P. Post-traumatic stress disorder. *BMJ*. 2015. Vol. 351. Art. h6161. URL: <https://doi.org/10.1136/bmj.h6161> (дата звернення: 26.09.2026).

71. Зимянський А., Скоропад В. Науково-психологічний аналіз дослідження проблеми посттравматичного стресового розладу в Україні. *Науковий вісник Ужгородського національного університету. Серія: Психологія*. 2025. № 1. С. 83–86. URL: <https://doi.org/10.32782/psy-visnyk/2025.1.15> (дата звернення: 19.09.2026).
72. World Health Organization. International Classification of Diseases, Eleventh Revision (ICD-11) : 6B40 Post traumatic stress disorder. Geneva : WHO, 2024. URL: <https://icd.who.int/browse/2024-01/mms/en#2070699808> (дата звернення: 26.09.2026).
73. Shalev A., Liberzon I., Marmar C. Post-Traumatic Stress Disorder. *New England Journal of Medicine*. 2017. Vol. 376, no. 25. P. 2459–2469. URL: <https://doi.org/10.1056/NEJMr1612499> (дата звернення: 26.09.2026).
74. Post-traumatic stress disorder: evolving conceptualization and evidence, and future research directions / C. R. Brewin et al. *World Psychiatry*. 2025. Vol. 24, no. 1. P. 52–80. URL: <https://doi.org/10.1002/wps.21269> (дата звернення: 19.05.2025).
75. Change in prevalence of post-traumatic stress disorder in the two years following trauma: a meta-analytic study / P. R. Diamond et al. *European Journal of Psychotraumatology*. 2022. Vol. 13, no. 1. URL: <https://doi.org/10.1080/20008198.2022.2066456> (дата звернення: 19.05.2025).
76. Using natural language processing to analyse text data in behavioural science / S. Feuerriegel et al. *Nature Reviews Psychology*. 2025. URL: <https://doi.org/10.1038/s44159-024-00392-z> (дата звернення: 19.05.2025).
77. Nguyen V. C., Lu N., Kane J. M., Birnbaum M. L., De Choudhury M. Cross-Platform Detection of Psychiatric Hospitalization via Social Media Data: Comparison Study. *JMIR Mental Health*. 2022. Vol. 9, No. 12. Art. e39747. DOI: <https://doi.org/10.2196/39747>
78. Daryanto, Windiarti I. S., Rintyarna B. S. Social Implications of Technological Advancements in Sentiment Analysis: A Literature Review on Potential and Consequences over the Next 20 Years. *The 8th Mechanical*

Engineering, Science and Technology International Conference. Basel Switzerland, 2025. P. 49. URL: <https://doi.org/10.3390/engproc2025084049> (дата звернення: 19.05.2025).

79. Kulmatov Z. Sentiment Analysis in Computational Linguistics: Bridging Technology and Human Emotion. *International Journal of Artificial Intelligence*. 2025. Vol. 1, no. 2. P. 1329–1332. URL: <https://inlibrary.uz/index.php/ijai/article/view/77262> (дата звернення: 22.05.2025).

80. Named Entity Recognition in Context: Edit_Dunhuang team Technical Report for Evahan2025 NER Competition / C. Brisson et al. *Proceedings of the Second Workshop on Ancient Language Processing*, The Albuquerque Convention Center, Laguna. Stroudsburg, PA, USA, 2025. P. 176–181. URL: <https://doi.org/10.18653/v1/2025.alp-1.22> (дата звернення: 31.03.2026).

81. Yates A., Cohan A., Goharian N. Depression and Self-Harm Risk Assessment in Online Forums. *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP 2017)*. Copenhagen, Denmark, 2017. P. 2968–2978. URL: <https://doi.org/10.18653/v1/D17-1322> (дата звернення: 26.09.2026).

82. Technology and Emotions: AI-Driven Software Prototyping for the Analysis of Emotional States and Early Detection of Risky Behaviors in University Students / A. C. Alves-Noreña et al. *Education Sciences*. 2025. Vol. 15, no. 3. P. 350. URL: <https://doi.org/10.3390/educsci15030350> (дата звернення: 22.05.2025).

83. Sending Messages in Social Networks / M. Cristani et al. *Agents and Multi-Agent Systems: Technologies and Applications 2018*. Cham, 2018. P. 123–133. URL: https://doi.org/10.1007/978-3-319-92031-3_12 (дата звернення: 22.05.2025).

84. Mental health amidst multiple crises: trends and sociodemographic risk factors in Austria's general population / E. Humer et al. *Frontiers in*

Psychiatry. 2025. Vol. 16. URL: <https://doi.org/10.3389/fpsy.2025.1534994> (дата звернення: 22.05.2025).

85. The Impact of Political Instability on Financial Anxiety, Career Orientation, Job Aspirations, Academic Performance, and Psychological Well-being among University Students / A. T. Khan et al. *Physical Education, Health and Social Sciences*. 2025. Vol. 3, Iss. 1. Pp. 161–171. URL: <https://journal-of-social-education.org/index.php/Jorunal/article/view/105> (дата звернення: 19.09.2026).

86. Mane S., Kundu S., Sharma R. A Survey on Online Aggression: Content Detection and Behavioural Analysis on Social Media Platforms. *ACM Computing Surveys*. 2025. URL: <https://doi.org/10.1145/3711125> (дата звернення: 22.05.2025).

87. Coppersmith G., Leary R., Crutchley P., Fine A. Natural Language Processing of Social Media as Screening for Suicide Risk. *Biomedical Informatics Insights*. 2018. Vol. 10. Art. 1178222618792860. URL: <https://doi.org/10.1177/1178222618792860> (дата звернення: 26.09.2026).

88. Analyzing two decades of media sentiments: NLP and deep learning insights into news bias and trends / S. Purohit et al. *Iran Journal of Computer Science*. 2025. URL: <https://doi.org/10.1007/s42044-025-00235-x> (дата звернення: 22.05.2025).

89. Dll5143A@NLU of Devanagari Script Languages 2025: Detection of Hate Speech and Targets Using Hierarchical Attention Network / A. Yadav et al. *Proceedings of the First Workshop on Challenges in Processing South Asian Languages (CHiPSAL 2025)*, Abu Dhabi, UAE, 19 January 2025. International Committee on Computational Linguistics, 2025. Pp. 278–288. URL: <https://aclanthology.org/2025.chipsal-1.31/> (дата звернення: 19.09.2026).

90. Acheampong F. A., Nunoo-Mensah H., Chen W. Transformer models for text-based emotion detection: a review of BERT-based approaches. *Artificial Intelligence Review*. 2021. URL: <https://doi.org/10.1007/s10462-021-09958-2> (дата звернення: 22.05.2025).

91. A Tool for Intelligent Text Analysis in Social Media / I. Smirnov et al. *Artificial Intelligence*. Cham, 2021. P. 232–247. URL: https://doi.org/10.1007/978-3-030-86855-0_16 (дата звернення: 22.05.2025).
92. Wahyuni E. D., Suryanto T. L. M., Arviani H. Deep Learning Multimodal Sarcasm Detection in Social Media Comments: The Role of Memes and Emojis. *Journal of Artificial Intelligence and Technology*. 2025. URL: <https://doi.org/10.37965/jait.2025.0699> (дата звернення: 22.05.2025).
93. Turcan E., McKeown K. Dreddit: A Reddit Dataset for Stress Analysis in Social Media. *Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis (LOUHI 2019)*. Hong Kong, 2019. P. 97–107. URL: <https://doi.org/10.18653/v1/D19-6213> (дата звернення: 26.09.2026).
94. CUET_HateShield@NLU of Devanagari Script Languages 2025: Transformer-Based Hate Speech Detection in Devanagari Script Languages / S. R. Aodhora et al. Proceedings of the First Workshop on Challenges in Processing South Asian Languages (CHiPSAL 2025), Abu Dhabi, UAE, 19 January 2025. *International Committee on Computational Linguistics*, 2025. Pp. 260–266. URL: <https://aclanthology.org/2025.chipsal-1.28/> (дата звернення: 19.09.2026).
95. Understanding and Discovering Deliberate Self-harm Content in Social Media / Y. Wang et al. *Proceedings of the 26th International Conference on World Wide Web (WWW 2017)*. Perth, Australia, 2017. P. 93–102. URL: <https://doi.org/10.1145/3038912.3052555> (дата звернення: 26.09.2026).
96. Cavicchio F. Resources for Emotion Detection: Lexicons and Annotated Datasets. *Synthesis Lectures on Human Language Technologies*. Cham, 2024. P. 33–43. URL: https://doi.org/10.1007/978-3-031-72047-5_3 (дата звернення: 22.05.2025).
97. Methods for Recognition of Frustration-Derived Reactions on Social Media / D. Devyatkin et al. *Artificial Intelligence*. Cham, 2021. P. 17–30. URL: https://doi.org/10.1007/978-3-030-86855-0_2 (дата звернення: 22.05.2025).

98. The Role of Artificial Intelligence in Optometric Diagnostics and Research: Deep Learning and Time-Series Forecasting Applications / L. F. F. M. Santos et al. *Technologies*. 2025. Vol. 13, no. 2. P. 77. URL: <https://doi.org/10.3390/technologies13020077> (дата звернення: 22.05.2025).

99. Choudhary J., Thakur M. Advances in AI for Women's Healthcare: A Review of Deep Learning and NLP Approaches. 2025 *3rd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, Bengaluru, India, 5–7 February 2025. 2025. P. 289–294. URL: <https://doi.org/10.1109/idciot64235.2025.10915173> (дата звернення: 19.09.2026).

100. Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence / V. Hassija et al. *Cognitive Computation*. 2023. URL: <https://doi.org/10.1007/s12559-023-10179-8> (дата звернення: 22.05.2025).

101. Explainable AI (XAI) for Energy Demand Forecasting / J. Nandha Gopal et al. *Neural Networks and Graph Models for Traffic and Energy Systems*. 2025. P. 115–138. URL: <https://doi.org/10.4018/979-8-3373-0290-4.ch005> (дата звернення: 22.05.2025).

102. Explainable AI (XAI) in Predictive Cloud Optimization: Cost, Workload and Performance / V. Viradia et al. *SoutheastCon 2025*, Concord, NC, USA, 22–30 March 2025. 2025. P. 1424–1429. URL: <https://doi.org/10.1109/southeastcon56624.2025.10971567> (дата звернення: 22.05.2025).

103. Hu X., Burke R., Liu X. Proposal for Workshop on Trust and Responsibility in Recommendation Systems at WSDM 2025. *WSDM '25: The Eighteenth ACM International Conference on Web Search and Data Mining*, Hannover Germany. New York, NY, USA, 2025. P. 1118–1119. URL: <https://doi.org/10.1145/3701551.3705707> (дата звернення: 22.05.2025).

104. Hedlund M., Persson E. Distribution of responsibility for AI development: expert views. *AI & SOCIETY*. 2025. URL: <https://doi.org/10.1007/s00146-024-02167-9> (дата звернення: 22.05.2025).

105. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI / A. Barredo Arrieta et al. *Information Fusion*. 2020. Vol. 58. P. 82–115. URL: <https://doi.org/10.1016/j.inffus.2019.12.012> (дата звернення: 26.09.2026).
106. Doshi-Velez F., Kim B. Towards A Rigorous Science of Interpretable Machine Learning. arXiv preprint arXiv:1702.08608. 2017. URL: <https://arxiv.org/abs/1702.08608> (дата звернення: 26.09.2026).
107. Rotari M., Kulahci M. Correlation to causality. *Quality Engineering*. 2025. Vol. 37, no. 1. P. 162–172. URL: <https://doi.org/10.1080/08982112.2024.2372489> (дата звернення: 22.05.2025).
108. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*. 2019. Vol. 1, no. 5. P. 206–215. URL: <https://doi.org/10.1038/s42256-019-0048-x> (дата звернення: 26.09.2026).
109. Post hoc calibration of medical segmentation models / A.-J. Rousseau et al. *Discover Applied Sciences*. 2025. Vol. 7, no. 3. URL: <https://doi.org/10.1007/s42452-025-06587-0> (дата звернення: 22.05.2025).
110. Ribeiro M. T., Singh S., Guestrin C. «Why Should I Trust You?»: Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2016)*. San Francisco, CA, USA, 2016. P. 1135–1144. URL: <https://doi.org/10.1145/2939672.2939778> (дата звернення: 26.09.2026).
111. Explainable Sentiment Analysis Utilizing Deep Learning Methods and LIME / K. Kyritsis et al. *Studies in Computational Intelligence*. Cham, 2025. P. 65–78. URL: https://doi.org/10.1007/978-3-031-78373-9_5 (дата звернення: 22.05.2025).
112. NER Explainability Framework: Utilizing LIME to Enhance Clarity and Robustness in Named Entity Recognition / M. Grundetjern et al. *Lecture Notes in Computer Science*. Cham, 2024. P. 3–15. URL: https://doi.org/10.1007/978-3-031-77915-2_1 (дата звернення: 22.05.2025).

113. Nie H., Wu X. A Dual-Channel Prediction-Interpretation Framework with Pre-Trained Language Models and SHAP Explainability. *Journal of Computer and Communications*. 2025. Vol. 13, no. 03. P. 116–137. URL: <https://doi.org/10.4236/jcc.2025.133009> (дата звернення: 22.05.2025).

114. Advancing eye disease detection: A comprehensive study on computer-aided diagnosis with vision transformers and SHAP explainability techniques / H. M. Balaha et al. *Biocybernetics and Biomedical Engineering*. 2025. Vol. 45, no. 1. P. 23–33. URL: <https://doi.org/10.1016/j.bbe.2024.11.005> (дата звернення: 22.05.2025).

115. Explainability to Image Captioning Models: An Improved Post-hoc Approach through Grad-CAM / H. Kumar et al. *2025 International Conference on Innovation in Computing and Engineering (ICE)*, Greater Noida, India, 28 February – 1 March 2025. 2025. P. 1–6. URL: <https://doi.org/10.1109/ice63309.2025.10984143> (дата звернення: 22.05.2025).

116. Explainable Artificial Intelligence with Integrated Gradients for the Detection of Adversarial Attacks on Text Classifiers / H. Moraliyage et al. *Applied System Innovation*. 2025. Vol. 8, no. 1. P. 17. URL: <https://doi.org/10.3390/asi8010017> (дата звернення: 22.05.2025).

117. Evaluating Attribution Methods for Explainable NLP with Transformers / V. Bartička et al. *Text, Speech, and Dialogue*. Cham, 2022. P. 3–15. URL: https://doi.org/10.1007/978-3-031-16270-1_1 (дата звернення: 22.05.2025).

118. Integrated gradients-based defense against adversarial word substitution attacks / M. Ahmed et al. *Neural Computing and Applications*. 2025. URL: <https://doi.org/10.1007/s00521-025-11135-3> (дата звернення: 22.05.2025).

119. Hoque E., Islam M. S. Natural Language Generation for Visualizations: State of the Art, Challenges and Future Directions. *Computer Graphics Forum*. 2024. URL: <https://doi.org/10.1111/cgf.15266> (дата звернення: 22.05.2025).

120. DeBERTa: Decoding-enhanced BERT with Disentangled Attention / P. He et al. *International Conference on Learning Representations (ICLR 2021)*. 2021. URL: <https://doi.org/10.48550/arXiv.2006.03654> (дата звернення: 19.09.2026).
121. SciDeBERTa: Learning DeBERTa for Science Technology Documents and Fine-Tuning Information Extraction Tasks / Y. Jeong et al. *IEEE Access*. 2022. Vol. 10. Pp. 60805–60813. URL: <https://doi.org/10.1109/ACCESS.2022.3180830> (дата звернення: 19.09.2026).
122. Kaggle. *Human Stress Prediction*. URL: <https://www.kaggle.com/datasets/kreeshrajani/human-stress-prediction/data>
123. Kaggle. *Aya ptsd*. URL: <https://www.kaggle.com/datasets/abdelrahmanahmed3/aya-ptsd>
124. Revolutionizing Mental Health Care through LangChain: A Journey with a Large Language Model / A. Singh et al. *2024 IEEE 14th Annual Computing and Communication Workshop and Conference (CCWC)*, Las Vegas, NV, USA, 8–10 January 2024. 2024. URL: <https://doi.org/10.1109/ccwc60891.2024.10427865> (дата звернення: 26.05.2025).
125. Deep Learning for Multi-Label Learning: A Comprehensive Survey / A. N. Tarekegn et al. *arXiv preprint arXiv:2401.16549*. 2024. URL: <https://arxiv.org/abs/2401.16549> (дата звернення: 19.09.2026).
126. From single label to extensive label set: A comprehensive review on Extreme multi-label classification / S. Gajjar et al. *AIP Conference Proceedings*. 2025. Vol. 3255, Iss. 1. Art. 020013. URL: <https://doi.org/10.1063/5.0254179> (дата звернення: 19.09.2026).
127. distilbert/distilbert-base-uncased / *Hugging Face*. 2023. URL: <https://huggingface.co/distilbert/distilbert-base-uncased> (дата звернення: 19.09.2026).
128. Method for analysis and formation of representative text datasets / O. Sobko et al. *CEUR Workshop Proceedings*. 2025. Vol. 3899. Pp. 84–98. (дата звернення: 19.09.2026).

129. Youden W. J. Index for rating diagnostic tests. *Cancer*. 1950. Vol. 3, no. 1. P. 32–35. URL: [https://doi.org/10.1002/1097-0142\(1950\)3:1<32::AID-CNCR2820030106>3.0.CO;2-3](https://doi.org/10.1002/1097-0142(1950)3:1<32::AID-CNCR2820030106>3.0.CO;2-3) (дата звернення: 01.10.2026).
130. Krak I., Zalutska O., Molchanova M., Mazurets O., Bahrii R., Sobko O., Barmak O. Abusive Speech Detection Method for Ukrainian Language Used Recurrent Neural Network. *CEUR Workshop Proceedings*, 2024, vol. 3688, pp. 16-28.
131. SamLowe/roberta-base-go_emotions · Hugging Face. *Hugging Face – The AI community building the future*. URL: https://huggingface.co/SamLowe/roberta-base-go_emotions (дата звернення: 31.03.2026).
132. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter / V. Sanh et al. *arXiv preprint arXiv:1910.01108*. 2019. URL: <https://arxiv.org/abs/1910.01108> (дата звернення: 19.09.2026).
133. COMSYS2023 Dataset / Kaggle. URL: <https://www.kaggle.com/datasets/kajimi/comsys2023> (дата звернення: 17.05.2025).
134. Depression Reddit Cleaned Dataset / Kaggle. URL: <https://www.kaggle.com/datasets/infamouscoder/depression-reddit-cleaned> (дата звернення: 17.05.2025).
135. Depression: Twitter Dataset + Feature Extraction. *Kaggle: The World's AI Proving Ground*. URL: <https://www.kaggle.com/datasets/infamouscoder/mental-health-social-media> (дата звернення: 31.03.2026).
136. Detecting cyberbullying using deep learning techniques: a pre-trained glove and focal loss technique / A. M. El Koshiry et al. *PeerJ Computer Science*. 2024. Vol. 10. P. e1961. URL: <https://doi.org/10.7717/peerj-cs.1961> (дата звернення: 19.09.2026).
137. Do we really need dice? The hidden region-size biases of segmentation losses / B. Liu et al. *Medical Image Analysis*. 2024. Vol. 91. Art.

103015. URL: <https://doi.org/10.1016/j.media.2023.103015> (дата звернення: 19.09.2026).

138. Mushava J., Murray M. Flexible loss functions for binary classification in Gradient-Boosted Decision Trees: an application to credit scoring. *Expert Systems with Applications*. 2024. Vol. 238. Art. 121876. URL: <https://doi.org/10.1016/j.eswa.2023.121876> (дата звернення: 19.09.2026).

139. Rainio O., Teuho J., Klén R. Evaluation metrics and statistical tests for machine learning. *Scientific Reports*. 2024. Vol. 14, no. 1. URL: <https://doi.org/10.1038/s41598-024-56706-x> (дата звернення: 19.09.2026).

140. microsoft/deberta-v3-small · Hugging Face. *Hugging Face – The AI community building the future*. URL: <https://huggingface.co/microsoft/deberta-v3-small> (дата звернення: 31.03.2026).

141. DeBERTaV3: Improving DeBERTa using ELECTRA-Style Pre-Training with Gradient-Disentangled Embedding Sharing / P. He et al. *arXiv preprint arXiv:2111.09543*. 2021. URL: <https://doi.org/10.48550/arXiv.2111.09543> (дата звернення: 19.09.2026).

142. Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters*. 2006. Vol. 27, no. 8. P. 861–874. URL: <https://doi.org/10.1016/j.patrec.2005.10.010> (дата звернення: 26.09.2026).

143. Pierse C. Transformers Interpret : model explainability for the Hugging Face transformers package. 2021. URL: <https://github.com/cdpierse/transformers-interpret> (дата звернення: 26.09.2026).

144. Vig J. A Multiscale Visualization of Attention in the Transformer Model. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations (ACL 2019)*. Florence, Italy, 2019. P. 37–42. URL: <https://doi.org/10.18653/v1/P19-3007> (дата звернення: 26.09.2026).

145. Enhancing suicidal behavior detection in EHRs: A multi-label NLP framework with transformer models and semantic retrieval-based annotation / K. Zandbiglari et al. *Journal of Biomedical Informatics*. 2025. Vol. 161. Art.

104755. URL: <https://doi.org/10.1016/j.jbi.2024.104755> (дата звернення: 19.09.2026).

146. All is attention for multi-label text classification / Z. Liu et al. *Knowledge and Information Systems*. 2024. URL: <https://doi.org/10.1007/s10115-024-02253-w> (дата звернення: 26.05.2025).

147. Research of multi-label text classification based on label attention and correlation networks / L. Yuan et al. *PLOS ONE*. 2024. Vol. 19, no. 9. P. e0311305. URL: <https://doi.org/10.1371/journal.pone.0311305> (дата звернення: 26.05.2025).

148. World Health Organization. Ethics and governance of artificial intelligence for health: WHO guidance. *Geneva : World Health Organization*, 2021. 150 p. URL: <https://www.who.int/publications/i/item/9789240029200> (дата звернення: 19.09.2026).

149. UNESCO. *Recommendation on the Ethics of Artificial Intelligence*. Paris : UNESCO, 2021. URL: <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics> (дата звернення: 19.09.2026).

150. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research*. Washington, D.C. : U.S. Department of Health and Human Services, 1979. URL: <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html> (дата звернення: 19.09.2026).

151. World Medical Association. *WMA Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Participants*. 2024. URL: <https://www.wma.net/policies-post/wma-declaration-of-helsinki/> (дата звернення: 19.09.2026).

152. Council for International Organizations of Medical Sciences. *International Ethical Guidelines for Health-related Research Involving Humans*. 4th ed. Geneva : CIOMS, 2016. 122 p. URL: <https://doi.org/10.56759/rgxl7405> (дата звернення: 19.09.2026).

153. Про захист персональних даних : Закон України від 01.06.2010 № 2297-VI. URL: <https://zakon.rada.gov.ua/go/2297-17> (дата звернення: 19.09.2026).

154. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 : General Data Protection Regulation. URL: <https://eur-lex.europa.eu/eli/reg/2016/679/oj> (дата звернення: 19.09.2026).

155. World Medical Association. *Declaration of Taipei: Ethical Considerations Regarding Health Databases and Biobanks*. Revised 2016. URL: <https://www.wma.net/what-we-do/medical-ethics/declaration-of-taipei/> (дата звернення: 19.09.2026).

156. Tabassi E. AI Risk Management Framework. Gaithersburg, MD : National Institute of Standards and Technology, 2023. URL: <https://doi.org/10.6028/nist.ai.100-1> (дата звернення: 19.09.2026).

157. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 : Artificial Intelligence Act. URL: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj> (дата звернення: 19.09.2026).

158. Method for post-traumatic stress disorder manifestation analyzing in text content / I. Krak, O. Ovcharuk, O. Mazurets, M. Molchanova, O. Sobko, O. Tyschenko, O. Barmak. *CEUR Workshop Proceedings*. 2025. Vol. 3976. P. 206–218. URL: <https://ceur-ws.org/Vol-3976/paper15.pdf> (дата звернення: 19.09.2026).

159. Transformer-based multilabel classification for identifying hidden psychological conditions in online posts / I. Krak, O. Mazurets, O. Ovcharuk, M. Molchanova, O. Barmak, L. Azarova. *CEUR Workshop Proceedings*. 2025. Vol. 4004. P. 86–97. URL: <https://ceur-ws.org/Vol-4004/paper8.pdf> (дата звернення: 19.09.2026).

160. Овчарук О. М., Мазурець О. В. Нейромережеве діагностування проявів ПТСР у текстовому контенті з використанням помилко-орієнтованого навчального набору даних. Вісник Хмельницького національного університету. Серія: Технічні науки. 2024. № 6, т. 1 (343). С.

195–200. URL: <https://doi.org/10.31891/2307-5732-2024-343-6-30> (дата звернення: 19.09.2026).

161. Овчарук О. М., Мазурець О. В. Нейромережевий метод діагностування психологічних розладів за аналізом повідомлень на основі роздільного підходу до класифікації. Вимірювальна та обчислювальна техніка в технологічних процесах. 2025. № 1. С. 210–216. URL: <https://doi.org/10.31891/2219-9365-2025-81-25> (дата звернення: 19.09.2026).

162. Овчарук О. М., Мазурець О. В. Превенція деструктивної поведінки за текстами користувачів з ПТСР методами обробки природної мови. Вісник Хмельницького національного університету. Серія: Технічні науки. 2025. Т. 359, № 6.2. С. 458–462. URL: <https://doi.org/10.31891/2307-5732-2025-359-136> (дата звернення: 19.09.2026).

163. Neural Network Assessment of Post-Traumatic Risk Behavior Using Social Media Posts Analysis / I. Krak, O. Ovcharuk, O. Mazurets, O. Sobko, M. Molchanova, O. Barmak. *CEUR Workshop Proceedings*. 2025. Vol. 4164. P. 141–151.

164. Овчарук О. М., Мазурець О. В. Метод аналізу проявів посттравматичного стресового розладу за нейромережевим аналізом текстів. Глушковські читання : матеріали XIII Всеукр. наук.-практ. конф. Київ, 2024. С. 218–221.

165. Мазурець О. В., Овчарук О. М. Діагностування проявів посттравматичного стресового розладу за нейромережевим аналізом текстового контенту. Нейромережні технології та їх застосування НМТіЗ-2024 : зб. наук. пр. XXIII Міжнар. наук. конф., 11–12 груд. 2024 р. Краматорськ–Тернопіль : ДДМА, 2024. С. 85–89.

166. Овчарук О. М., Мазурець О. В. Підхід до оцінювання поведінкових ризиків користувачів з ПТСР за цифровим профілем засобами NLP. Математичне та програмне забезпечення інтелектуальних систем (МПЗІС-2025) : тези доп. XXIII Міжнар. наук.-практ. конф., 19–21 листоп. 2025 р. Дніпро : ДНУ, 2025. С. 241–242.

167. Комп'ютерна програма «Інтелектуальна інформаційна система для аналізу прояву посттравматичного стресового розладу в текстовій інформації» : свідоцтво про реєстрацію авторського права на твір № 135968 Україна / О. В. Мазурець, О. М. Овчарук. Дата реєстрації: 08.05.2025.

168. Комп'ютерна програма «Інтелектуальна інформаційна система для виявлення психічних розладів за аналізом текстової інформації» : свідоцтво про реєстрацію авторського права на твір № 135967 Україна / О. В. Мазурець, О. М. Овчарук. Дата реєстрації: 08.05.2025.

169. Mazurets O. V., Ovcharuk O. M. Methodology for Identifying Post-Traumatic Stress Disorder Indicators in Text Data. Information Control Systems and Intelligent Technologies. Advances and Applications : monograph. Liha-Pres, 2025. P. 43–58. URL: <https://doi.org/10.36059/978-966-397-538-2-3> (дата звернення: 19.09.2026).

ДОДАТОК А.
СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА

*Статті у наукових виданнях, включених до Переліку наукових фахових
видань України:*

1. Овчарук О.М., Мазурець О.В. Нейромережеве діагностування проявів ПТСР у текстовому контенті з використанням помилко-орієнтованого навчального набору даних. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2024. №6, Т.1 (343). С. 195-200. <http://doi.org/10.31891/2307-5732-2024-343-6-30>.

2. Овчарук О.М., Мазурець О.В. Нейромережевий метод діагностування психологічних розладів за аналізом повідомлень на основі роздільного підходу до класифікації. Міжнародний науково-технічний журнал «Вимірювальна та обчислювальна техніка в технологічних процесах». 2025. № 1. С. 210-216. <https://doi.org/10.31891/2219-9365-2025-81-25>.

3. Овчарук О.М., Мазурець О.В. Превенція деструктивної поведінки за текстами користувачів з ПТСР методами обробки природної мови. Вісник Хмельницького національного університету. Серія: Технічні науки. 2025. №6, Т.2 (359). С. 458–462. <https://doi.org/10.31891/2307-5732-2025-359-136>.

Праці, які засвідчують апробацію матеріалів дисертації:

4. Krak I., Ovcharuk O., Mazurets O., Molchanova M., Sobko O., Tyschenko O., Barmak O. Method for post-traumatic stress disorder manifestation analyzing in text content. CEUR Workshop Proceedings. 2025. Vol. 3976. P. 206-218. URL: <https://ceur-ws.org/Vol-3976/paper15.pdf>

5. Krak I., Mazurets O., Ovcharuk O., Molchanova M., Barmak O., Azarova L. Transformer-based multilabel classification for identifying hidden psychological conditions in online posts. CEUR Workshop Proceedings. 2025. Vol. 4004. P. 86-97. URL: <https://ceur-ws.org/Vol-4004/paper8.pdf>

6. Krak I., Ovcharuk O., Mazurets O., Sobko O., Molchanova M., Barmak O. Neural Network Assessment of Post-Traumatic Risk Behavior Using Social

Media Posts Analysis. CEUR Workshop Proceedings. 2025. Vol. 4164. P. 141-151. URL: <https://ceur-ws.org/Vol-4164/paper9.pdf>

7. Овчарук О.М., Мазурець О.В. Метод аналізу проявів посттравматичного стресового розладу за нейромережовим аналізом текстів. Матеріали XIII Всеукраїнської науково-практичної конференції «Глушковські читання». 6 грудня 2024. Київ, 2024. С. 218-221.

8. Мазурець О.В., Овчарук О.М. Діагностування проявів посттравматичного стресового розладу за нейромережовим аналізом текстового контенту. Збірник наукових праць XXIII Міжнародної наукової конференції «Нейромережні технології та їх застосування НМТіЗ-2024». 11-12 грудня 2024. Краматорськ-Тернопіль, ДДМА. 2024. С. 85-89.

9. Овчарук О.М., Мазурець О.В. Підхід до оцінювання поведінкових ризиків користувачів з ПТСР за цифровим профілем засобами NLP. Тези доповідей XXIII Міжнародної науково-практичної конференції «Математичне та програмне забезпечення інтелектуальних систем МПЗІС-2025». 19-21 листопада 2025. Дніпро: ДНУ, 2025. С. 241-242.

Публікації, які додатково відображають наукові результати дисертації:

10. Мазурець О.В., Овчарук О.М. Комп'ютерна програма «Інтелектуальна інформаційна система для аналізу прояву посттравматичного стресового розладу в текстовій інформації»: свідоцтво про реєстрацію авторського права № 135968 від 08.05.2025.

11. Мазурець О.В., Овчарук О.М. Комп'ютерна програма «Інтелектуальна інформаційна система для виявлення психічних розладів за аналізом текстової інформації»: свідоцтво про реєстрацію авторського права № 135967 від 08.05.2025.

12. Mazurets O.V., Ovcharuk O.M. Methodology for Identifying Post-Traumatic Stress Disorder Indicators in Text Data. Information Control Systems and Intelligent Technologies. Advances and Applications. Monograph. Liha-Pres. 2025. P. 43-58. ISBN 978-966-397-538-2. <https://doi.org/10.36059/978-966-397-538-2-3>.

ДОДАТОК Б.
АКТИ ВПРОВАДЖЕННЯ

ДОВІДКА
про впровадження результатів дисертаційної роботи
Овчарука Олександра Миколайовича
«Методи аналізу психічного стану в громадян із посттравматичним стресовим
розладом за інтелектуальним аналізом користувацького контенту»

Наступні результати дисертаційної роботи Овчарука Олександра Миколайовича «Методи аналізу психічного стану в громадян із посттравматичним стресовим розладом за інтелектуальним аналізом користувацького контенту» були впроваджені у виробничу діяльність ПП «Авіві»: метод аналізу прояву посттравматичного стресового розладу в користувацькому контенті, метод виявлення психічних розладів за аналізом користувацького контенту та метод розширення множини ознак користувацького контенту для попередження соціально небезпечної поведінки громадян із посттравматичним стресовим розладом.

Зазначені результати були використані при розробці інтелектуальних компонентів аналітичних сервісів, орієнтованих на аналіз текстових звернень і повідомлень користувачів корпоративних систем та чат-ботів. Запропонований метод аналізу прояву посттравматичного стресового розладу на основі контекстно-орієнтованої нейромережевої моделі DeBERTa та спеціалізованого навчального набору даних було інтегровано у процес обробки текстових даних для виявлення повідомлень з ознаками посттравматичного стресового розладу. Метод виявлення психічних розладів, реалізований як множина незалежних бінарних трансформерних моделей з індивідуальними порогами, дав змогу одночасно ідентифікувати ознаки кількох супутніх психічних станів в одному повідомленні. Метод розширення множини ознак користувацького контенту було застосовано для формування розширеного ознакового профілю користувача та часового індексу ризику з поясненням результатів аналізу, що підвищило прозорість функціонування системи та спростило інтерпретацію результатів для кінцевих користувачів.

Практична апробація та дослідна експлуатація розроблених рішень підтвердили їх ефективність у задачах автоматизованого аналізу текстового контенту, зокрема з точки зору точності виявлення ознак посттравматичного стресового розладу і супутніх психічних розладів та інформативності отриманих пояснень.

29.04.2026

Директор ПП «Авіві»

Вісвіт

Аскеров В. В.



ТзОВ «Стфалкон»
Україна, м. Хмельницький,
вул. Зарічанська 3/2
Код ЄДРПОУ 41337336

Довідка

про впровадження результатів дисертаційної роботи Овчарука Олександра Миколайовича
«Методи аналізу психічного стану в громадян із посттравматичним стресовим розладом за
інтелектуальним аналізом користувацького контенту»
у виробничу діяльність ТОВ «Стфалкон»

Результати дисертаційної роботи Овчарука Олександра Миколайовича на тему «Методи аналізу психічного стану в громадян із посттравматичним стресовим розладом за інтелектуальним аналізом користувацького контенту» були впроваджені у виробничу діяльність ТОВ «Стфалкон» при розробці інтелектуальних модулів аналізу користувацького текстового контенту для вебзастосунків і мобільних застосунків. До переліку впроваджених результатів належать: метод аналізу прояву посттравматичного стресового розладу в користувацькому контенті; метод виявлення психічних розладів за аналізом користувацького контенту; метод розширення множини ознак користувацького контенту для попередження соціально небезпечної поведінки громадян із посттравматичним стресовим розладом.

Зазначені результати використано при проектуванні та програмній реалізації серверних компонентів, що виконують нейромережевий аналіз текстових повідомлень користувачів. Метод аналізу прояву ПТСР застосовано для автоматизованого виявлення повідомлень з ознаками посттравматичного стресового розладу з урахуванням контексту висловлювань; метод виявлення психічних розладів – для багатоміткової ідентифікації ознак психічних станів; метод розширення множини ознак – для побудови розширеного профілю користувача, що поєднує лінгвістичні, емоційно-семантичні та поведінкові показники його цифрової активності, а також скринінгового оцінювання ризику соціально небезпечної поведінки. Засоби візуальної інтерпретації рішень нейромережевих моделей використано для пояснення фахівцям результатів.

Впровадження та дослідна експлуатація розроблених рішень підтвердили їх працездатність і практичну цінність для створення програмних продуктів у сфері цифрового психологічного моніторингу та підтримки психічного здоров'я, зокрема, скоротили трудомісткість розроблення відповідних аналітичних модулів і підвищили якість автоматизованого аналізу текстового контенту.

19.05.2026

Директор ТОВ «Стфалкон»



Степан ТАНАСІЙЧУК

Довідка
про впровадження результатів дисертаційної роботи
Овчарука Олександра Миколайовича
«Методи аналізу психічного стану в громадян із посттравматичним стресовим розладом за інтелектуальним аналізом користувацького контенту»

Результати дисертаційної роботи Овчарука Олександра Миколайовича на тему «Методи аналізу психічного стану в громадян із посттравматичним стресовим розладом за інтелектуальним аналізом користувацького контенту» були використані в межах діяльності громадської організації «ІТ-кластер міста Хмельницького» при розробці програмних складових сервісу моніторингу психоемоційного стану та психологічної підтримки учасників цифрової спільноти. Зокрема, були впроваджені результати дисертаційної роботи: метод аналізу прояву посттравматичного стресового розладу в користувацькому контенті, який завдяки фокусуванню на контекстних залежностях, специфічних для ПТСР, та використанню спеціалізованого навчального набору даних дає змогу відмежовувати прояви ПТСР від проявів інших психічних розладів; метод виявлення психічних розладів за аналізом користувацького контенту, який забезпечує ідентифікацію одразу кількох психічних розладів, а не лише домінуючого; метод розширення множини ознак користувацького контенту для попередження соціально небезпечної поведінки громадян із посттравматичним стресовим розладом, який враховує не лише лінгвістичні характеристики, а й емоційно-семантичні та поведінкові патерни цифрової активності користувачів.

Розроблене на основі зазначених результатів програмне забезпечення забезпечило автоматизований аналіз текстових повідомлень користувачів. Зокрема, програмний модуль формує інтегральну оцінку ймовірності наявності ознак ПТСР у текстах, виявляє ознаки супутніх психічних станів, буде часовий індекс ризику соціально небезпечної поведінки, а також надає пояснення отриманих висновків шляхом візуальної інтерпретації результатів нейромережевого аналізу. За результатом впровадження в діяльність громадської організації «ІТ-кластер міста Хмельницького», це дозволило здійснювати скринінговий моніторинг психоемоційного стану користувачів у цифровому середовищі та своєчасно виявляти осіб, які потребують психологічної підтримки.

Заступник голови
ГО «ІТ-кластер міста Хмельницького»

Сергій ЯЦИШЕН

21.05.2026р.





«ЗАТВЕРДЖУЮ»

Проректор з наукової роботи
Хмельницького національного університету

 Олег СИНЮК

7 » вересня 2026 р.

АКТ

про впровадження в навчальний процес результатів досліджень
аспіранта Овчарука Олександра Миколайовича за темою дослідження
«Методи аналізу психічного стану в громадян із посттравматичним
стресовим розладом за інтелектуальним аналізом користувацького контенту»

Результати дисертаційної роботи Овчарука Олександра Миколайовича, а саме, метод аналізу прояву посттравматичного стресового розладу в користувацькому контенті, метод виявлення психічних розладів за аналізом користувацького контенту та метод розширення множини ознак користувацького контенту для попередження соціально небезпечної поведінки громадян із посттравматичним стресовим розладом, використовуються в навчальному процесі при викладанні дисципліни бакалаврського рівня «Методи та системи штучного інтелекту» та дисципліни магістерського рівня «Моделі та методи текстової аналітики», виконанні кваліфікаційних робіт здобувачів бакалаврського та магістерського рівнів вищої освіти спеціальності «Комп'ютерні науки».

Акт обговорений і схвалений на засіданні кафедри комп'ютерних наук
(протокол № 3 від 07 вересня 2026 р.).

Зав. каф. комп'ютерних наук,
канд. техн. наук, доцент



Олександр МАЗУРЕЦЬ

Секретар каф. комп'ютерних наук,
ст. викладач



Тетяна СКРИПНИК

ДОДАТОК В. ПРОГРАМНІ КОДИ

Вихідний код, використаний у дослідженні, доступний у репозиторії GitHub:
<https://github.com/oovcharuk/MentalHealthTextClassifier/tree/identifying-risky-behavior> (дата звернення: 31.03.2026). На рис. В.1 наведено знімок екрана репозиторію.

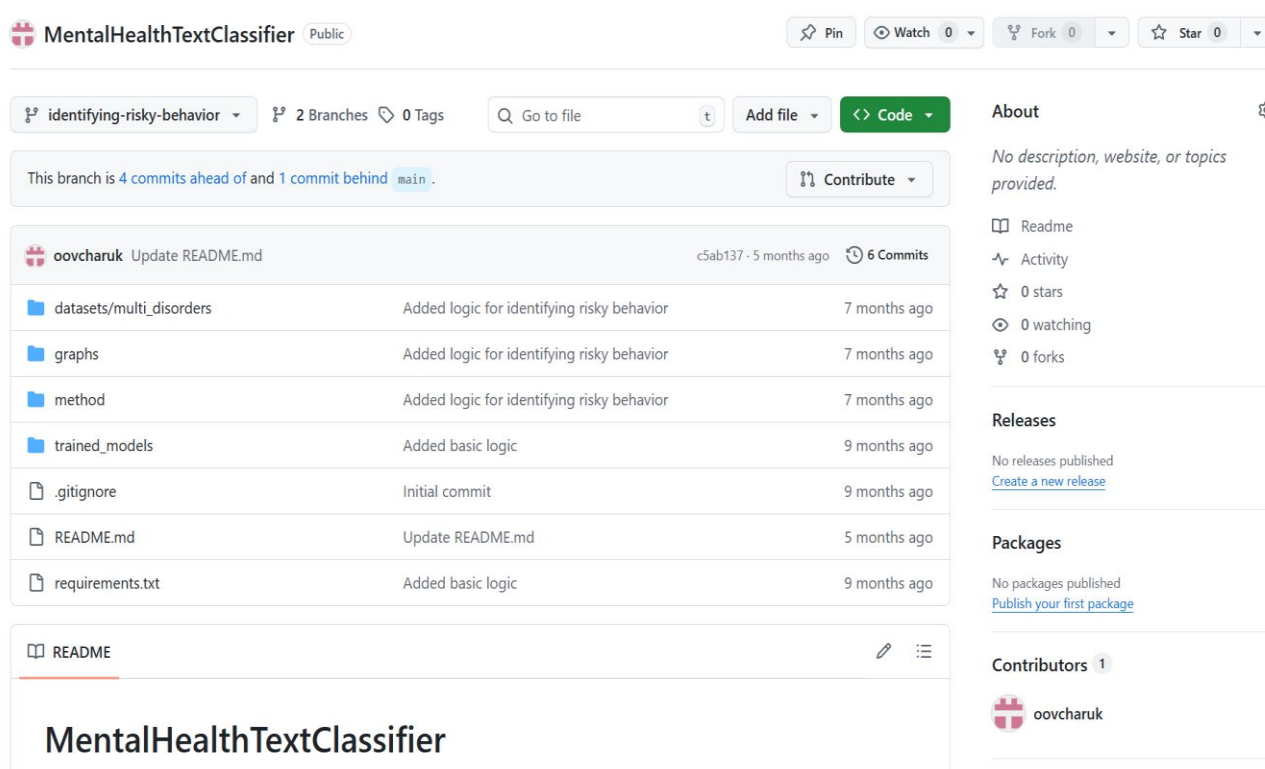


Рис. В.1 – Головна сторінка репозиторію GitHub з програмним кодом

Структура репозиторію наступна:

- основні програмні модулі методів (method). Містять реалізацію алгоритмів обробки тексту, навчання моделі та класифікації повідомлень для виявлення ознак психічних розладів і ризикованої поведінки;
- модулі для роботи з наборами даних (datasets/multi_disorders). Містять дані, що використовуються для навчання, валідації та аналізу психічних розладів і соціально-небезпечної поведінки громадян;

– модулі для побудови графіків і візуалізації результатів (graphs).
Призначені для графічного подання результатів експериментів та інтерпретації отриманих рішень;

– модулі для зберігання навчених моделей (trained_models).
Використовується для зберігання навчених моделей, що використовуються для тестування, оцінювання та відтворення результатів дослідження.

ДОДАТОК Г.
АВТОРСЬКІ СВДОЦТВА

УКРАЇНА



СВДОЦТВО

про реєстрацію авторського права на твір

№ 135968

Комп'ютерна програма «Інтелектуальна інформаційна система для аналізу прояву посттравматичного стресового розладу в текстовій інформації»

(вид, назва твору)

Автор (співавтори) **Мазурець Олександр Вікторович, Овчарук Олександр Миколайович**

(прізвище, ім'я, по батькові (за наявності), псевдонім (за наявності))

Авторські майнові права належать спільно **Мазурець Олександр Вікторович, вул. _____, м. Хмельницький, 29027; Овчарук Олександр Миколайович, вул. _____, м. Хмельницький, 29019**

(прізвище, ім'я, по батькові (за наявності) фізичної особи / найменування юридичної особи, адреса)

Дата реєстрації 8 травня 2025 р.

Директор Державної організації
«Український національний
офіс інтелектуальної власності
та інновацій»


Олена ОРЛЮК



УКРАЇНА



СВІДОЦТВО

про реєстрацію авторського права на твір

№ 135967

Комп'ютерна програма «Інтелектуальна інформаційна система для виявлення психічних розладів за аналізом текстової інформації»

(вид, назва твору)

Автор (співавтори) Мазурець Олександр Вікторович, Овчарук Олександр Миколайович

(прізвище, ім'я, по батькові (за наявності), псевдонім (за наявності))

Авторські майнові права належать спільно Мазурець Олександр Вікторович, вул. [redacted], м. Хмельницький, 29027; Овчарук Олександр Миколайович, [redacted], м. Хмельницький, 29019

(прізвище, ім'я, по батькові (за наявності) фізичної особи / найменування юридичної особи, адреса)

Дата реєстрації 8 травня 2025 р.

Директор Державної організації
«Український національний
офіс інтелектуальної власності
та інновацій»

Олена ОРЛЮК

