Голові разової спеціалізованої
вченої ради PhD 17283
Хмельницького національного університету
доктору технічних наук, професору
Едуарду МАНЗЮКУ

ВІДГУК

офіційного опонента на дисертаційну роботу
Овчарука Олександра Миколайовича
за темою «Методи аналізу психічного стану в громадян із посттравматичним стресовим розладом за інтелектуальним аналізом користувацького контенту»,
подану на здобуття наукового ступеня доктора філософії
з галузі знань 12 «Інформаційні технології»
за спеціальністю 122 «Комп'ютерні науки»

Актуальність теми та зв'язок з науковими планами та програмами.

Дисертаційна робота Овчарука Олександра Миколайовича присвячена розробленню методів аналізу психічного стану громадян із посттравматичним стресовим розладом за інтелектуальним аналізом користувацького контенту. Обрана тематика є актуальною, оскільки в умовах тривалої повномасштабної війни кількість осіб, які потребуватимуть психологічної підтримки, за наведеними у роботі оцінками сягає 40–50 % населення України, а значна частина постраждалих не звертається по фахову допомогу. За таких умов користувацький контент соціальних мереж стає одним із небагатьох доступних джерел інформації про психоемоційний стан людини, яке можна аналізувати без прямого втручання в її особистий простір.

Наукова своєчасність роботи визначається міждисциплінарним характером поставленої задачі, що перебуває на перетині обробки природної мови, глибокого навчання, багатоміткової класифікації та клінічної психології. Важливим є те, що здобувач не зводить задачу до чергової бінарної класифікації «ПТСР – не ПТСР», а розглядає три взаємопов'язані рівні аналізу: виявлення ознак ПТСР у повідомленні, виявлення супутніх психічних розладів без примусового вибору одного домінуючого стану та перехід від рішень щодо окремих повідомлень до скринінгового часового індексу ризику на рівні користувача. Суттєва складність досліджуваної задачі полягає в тому, що мовні маркери ПТСР, тривожних і депресивних станів значною мірою перетинаються, а наявні відкриті набори даних розмічено за тематикою онлайн-спільнот, а не за

клінічно підтвердженими діагнозами. Тому запропоноване в дисертації спрямування на контрольоване відмежування ознак ПТСР від тематично суміжних станів шляхом спеціального формування нецільового класу навчальної вибірки є методологічно виправданим та науково значущим.

Доцільним є також обраний здобувачем підхід до оцінювання ризику соціально небезпечної поведінки не за окремим дописом, а за часовою динамікою профілю користувача в рухомому вікні. Така постановка відповідає вимогам до скринінгових, а не діагностичних систем і узгоджується з Цільовою моделлю 2.0 системи психічного здоров'я та психосоціальної підтримки в Україні. Соціальна значущість дослідження посилюється тим, що його результати орієнтовані на підтримку фахівців і соціальних служб у ранньому виявленні осіб, які потребують допомоги, а не на автономне ухвалення рішень алгоритмом.

Зв'язок роботи з науковими програмами, планами і темами підтверджується тим, що дисертацію виконано в межах держбюджетних тем Хмельницького національного університету, у яких здобувач був виконавцем: «Розробка методів і 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).

Аналіз змісту дисертації та основні результати роботи.

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.

Варто зазначити, що здобувач чітко операціоналізує мету: під точністю розуміються показники класифікації на рівні повідомлення, а під розширенням можливостей оцінювання ризиків – перехід до скринінгового часового індексу ризику на рівні користувача (с. 24, 52–53). Таке уточнення дає змогу однозначно перевірити досягнення мети за результатами розділу 4.

Загалом дисертаційна робота коректно структурована відповідно до поставленої мети і завдань дослідження. Робота написана на достатньому мовно-стилістичному рівні. Її зміст дає змогу скласти цілісне уявлення про основні наукові положення, запропоновані здобувачем, їх практичне значення, а також результати експериментального дослідження розроблених методів.

Окремо варто відзначити наявність п. 4.6 «Обмеження запропонованих методів та використаних наборів даних» (с. 140–148), у якому здобувач самостійно і досить детально систематизує обмеження складу наборів даних, протоколу оцінювання та зовнішньої валідності результатів, що є нетиповим для дисертацій такого рівня і заслуговує на схвалення.

У *першому розділі* дисертації проведено аналіз методів, засобів та технологій для автоматизованого виявлення ПТСР у користувацькому текстовому контенті. Розглянуто сучасний стан інтелектуального аналізу текстів для виявлення ПТСР та інших психічних розладів, проаналізовано урядові програми і нормативні документи у сфері психічного здоров'я, систематизовано ознаки психічних розладів та соціально небезпечної поведінки у текстовому контенті, а також підходи до пояснюваності нейромережевих рішень. За результатами аналізу сформульовано три невирішені проблеми та відповідно мету і 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. Заслуговує на увагу п. 2.5, у якому формально показано, що комплексний аналіз не є простою композицією операторів, визначено повний та умовно-послідовний режими функціонування комплексу, доведено тотожність підсумкового індексу ризику в обох режимах.

У *третьому розділі* наведено алгоритмічну реалізацію запропонованих методів: нейромережеву модель DeBERTa-v3-small для виявлення ПТСР, множину з п'яти незалежних бінарних моделей DistilBERT з індивідуальними порогами за індексом Юдена для багатоміткового виявлення психічних розладів, а також алгоритм обчислення часового індексу ризику. Описано формування спеціалізованих наборів даних, метрики оцінювання та функції втрат, дизайн трьох сценаріїв експериментального дослідження та етичні аспекти й межі автоматизованих рішень.

У *четвертому розділі* наведено результати експериментального дослідження. Для методу аналізу прояву ПТСР на валідаційній вибірці з 440 записів отримано Accuracy 0,934, Precision 0,948, F1-score 0,933, чутливість 0,918 і специфічність 0,950. Для методу виявлення психічних розладів показано, що перехід від багатокласової моделі (Macro F1-score 0,862) до

множини бінарних класифікаторів з адаптивними порогами забезпечує Macro Accuracy 0,9316, Macro Precision 0,9338, Macro Recall 0,9296 та Macro F1-score 0,9296. Для методу розширення множини ознак на профілях користувачів набору даних із часовими мітками показано, що індекс ризику розрізняє профілі з епізодичними сплесками, без коливань та зі стійко високим рівнем, а його складові пояснювані завдяки адитивності. Важливим є те, що здобувач на конкретних прикладах виявляє джерела хибнопозитивних рішень – сарказм та ігровий контекст – і коректно інтерпретує їх як обмеження методів рівня повідомлення.

У висновках дисертації подано узагальнення отриманих результатів за кожним із поставлених завдань, окремо сформульовано межі застосування результатів та напрями подальших досліджень. Список використаних джерел відповідає предметній області та охоплює як сучасні праці з обробки природної мови й трансформерних моделей, так і клінічні настанови та нормативні документи у сфері психічного здоров'я. Додатки містять список публікацій, акти впровадження, відомості про програмні коди з посиланням на відкритий репозиторій та авторські свідоцтва.

Основні результати роботи:

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

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

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

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

5. Розроблено програмну систему інтелектуального аналізу користувацького контенту та експериментально досліджено ефективність запропонованих методів.

Наукова новизна, оцінка обґрунтованості наукових положень дисертації та їх достовірності.

Загалом наукові положення, висновки та рекомендації, сформульовані у дисертаційній роботі, є достатньо обґрунтованими. Їх достовірність

підтверджується коректною постановкою задач, формалізацією методів, застосуванням адекватного методичного апарату, експериментальною перевіркою запропонованих методів з усередненням результатів за чотирма випадковими розбиттями для другого методу, програмною реалізацією з відкритим кодом, апробацією результатів у наукових публікаціях та наявністю впроваджень. Сукупність отриманих результатів дає підстави вважати, що дисертаційна робота містить нові науково обґрунтовані положення, які мають значення для розвитку методів інтелектуального аналізу текстових даних у задачах моніторингу психічного здоров'я.

Наукова новизна отриманих результатів полягає в наступному:

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 порогами 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.

Теоретичне та практичне значення одержаних результатів.

Наукова цінність дисертаційної роботи полягає в тому, що здобувачем обґрунтовано, формалізовано та експериментально підтверджено підхід до аналізу психічного стану громадян із ПТСР, у якому методи рівня повідомлення та рівня користувача узгоджено в єдиний комплекс з визначеними інтерфейсами, умовною активацією поглиблених етапів, інваріантністю часової оцінки до калібрування моделей та прогнозованим поширенням помилок. На відміну від підходів, орієнтованих на виявлення

одного домінуючого розладу або на оцінювання поведінкових ризиків за окремими лінгвістичними ознаками, у роботі ризик розглядається як часовий процес на рівні користувача, а не як властивість окремого повідомлення, що дозволяє відмежовувати поодинокі, зокрема хибнопозитивні, рішення від стійких тенденцій.

Практичне значення отриманих результатів полягає у програмній реалізації розроблених методів у вигляді системи інтелектуального аналізу користувацького контенту, компоненти якої захищено двома свідоцтвами про реєстрацію авторського права, а програмний код розміщено у відкритому репозиторії GitHub. Розроблене програмне забезпечення дає змогу формувати модельну оцінку прояву ознак ПТСР у текстах, виявляти супутні психічні стани, будувати часовий ряд індексу ризику та візуалізувати результати роботи моделей, що підвищує не лише точність, а й інтерпретованість моніторингу стану користувача.

Розроблені в роботі методи дозволили досягти поставленої мети дисертаційного дослідження. Експериментальні результати підтверджують працездатність запропонованого підходу: для методу аналізу прояву ПТСР отримано Accuracy 0,934, Precision 0,948 та F1-score 0,933; для методу виявлення психічних розладів – Macro F1-score 0,9296 проти 0,862 для багатокласової моделі на тих самих класах.

Практична затребуваність результатів підтверджується їх упровадженням в освітній процес Хмельницького національного університету, а також апробацією розробок у діяльності ГО «ІТ Кластер м. Хмельницького», ПП «Авіві» та ТОВ «Стфалкон». Результати також пов'язані з виконанням держбюджетних тем Хмельницького національного університету.

Оцінка змісту дисертації, її завершеність та дотримання принципів академічної доброчесності.

За своїм змістом дисертаційна робота здобувача Овчарука Олександра Миколайовича повністю відповідає Стандарту вищої освіти зі спеціальності 122 Комп'ютерні науки для третього (освітньо-наукового) рівня вищої освіти. Дисертаційна робота є завершеною науковою працею і свідчить про наявність особистого внеску здобувача у науковий напрям комп'ютерних наук. Розглянувши результати перевірки дисертаційної роботи, можна зробити висновок, що дисертаційна робота Овчарука О. М. є результатом самостійних досліджень здобувача і не містить ознак академічного плагіату, фабрикації, фальсифікації або необґрунтованих текстових запозичень. Використані результати і тексти інших авторів мають належні посилання на відповідне

джерело. На титульному аркуші дисертації здобувач окремо засвідчує, що всі частини тексту, під час написання яких використовувались технології штучного інтелекту, перевірені та відредаговані особисто. Дисертація написана з дотриманням принципів академічної доброчесності.

Оприлюднення результатів дисертаційної роботи.

За результатами досліджень дисертації опубліковано 12 наукових праць, з-поміж яких: 3 статті у наукових фахових виданнях України категорії Б; 6 публікацій, які засвідчують апробацію матеріалів дисертації, з яких 3 опубліковано у збірниках CEUR Workshop Proceedings, що індексуються у наукометричній базі Scopus; 3 публікації, які додатково відображають наукові результати дисертації, зокрема розділ у колективній монографії та 2 свідоцтва про реєстрацію авторського права на твір. Тобто здобувач має не менше 3 публікацій, у яких викладені основні результати дисертації, чого достатньо згідно з п. 8 Постанови Кабінету Міністрів України від 12 січня 2022 р. № 44. У спільних публікаціях чітко зазначено особистий внесок дисертанта та внесок кожного зі співавторів. У всіх публікаціях здобувачем дотримано принципів академічної доброчесності.

Зауваження та побажання.

При позитивній оцінці дисертаційної роботи вважаю за необхідне зробити такі зауваження:

1. Внесок складових першого методу. Наукова новизна першого методу (с. 25) формулюється як поєднання двох складових: спеціалізованого формування нецільового класу навчальної вибірки та контекстно-позиційного представлення тексту моделлю DeBERTa. Водночас на сторінці 146 здобувач сам зазначає, що 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 окремо не оцінено. Фактично отримані значення Accuracy 0,934 та F1-score 0,933 (с. 114) характеризують метод у цілому, а твердження про те, що саме включення текстів з проявами інших розладів до нецільового класу «дало змогу підвищити точність виявлення ознак ПТСР», залишається експериментально не відокремленим від ефекту самої архітектури. Такий абляційний експеримент не потребує нових даних і суттєво посилив би доказовість положення наукової новизни.

2. Протокол оцінювання першого методу. На сторінці 84 зазначено, що набір даних поділено на навчальну та валідаційну вибірки у співвідношенні 80 % до 20 %, а на сторінці 146 – що ця сама валідаційна вибірка

використовувалась і для контролю навчання, окремої тестової вибірки не виділяли, а результат одержано за один запуск навчання. Крім того, 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 одночасно до навчальної і валідаційної вибірок. Здобувач сам визнає, що наведені значення можуть бути дещо завищеними. Було б доцільно виконати групове розбиття за post_id та навести розкид метрик за кількома запусками, як це зроблено для другого методу (ΔF1 у табл. 4.4, 4.7, 4.8).

3. На сторінці 115 вказано, що ROC-криву на рис. 4.3 побудовано за бінарними рішеннями моделі, тому вона має одну точку зламу, а значення 0,872 не є ранговою ROC-AUC і у підсумкових результатах не використовується. Тобто для основної моделі роботи рангова ROC-AUC не обчислена, хоча для моделі з квантовим шаром (табл. 4.1) та для всіх п'яти бінарних моделей другого методу (рис. 4.11–4.15) такі криві наведено. Обчислення AUC за ймовірностями позитивного класу не потребує повторного навчання і дало б змогу коректно зіставити модель DeBERTa з іншими моделями роботи.

4. У рукописі використано багаторівневу систему порогів, однак їх значення задано з якісних міркувань без аналізу чутливості:

А) поріг рішення для модельної оцінки прояву ПТСР $\tau = 0,5$ (сторінка 93) взято як стандартне значення, але не показано, що саме це значення є оптимальним для задачі;

Б) пороги 0,30 і 0,50 та коефіцієнти 0,2 і 0,3 у (2.17)–(2.18) обґрунтовано лише правилами «понад третину» і «понад половину» (сторінка 70), а на сторінці 147 прямо зазначено, що аналіз чутливості індексу до цих параметрів і до ширини вікна $W = 30$ днів не виконували;

В) пороги бінарних моделей за індексом Юдена (від 0,501 до 0,8978, сторінки 130–132) визначено за передбаченнями на навчальній вибірці (крок 8 алгоритму 3.5, сторінка 90), а не на відкладеній підвибірці, що може давати оптимістичну оцінку; при цьому варіанти з адаптивним порогом і без нього відрізняються ще й розміром пачки та кількістю епох, тому приріст Macro F1-score з 0,925 до 0,9296 не можна приписати лише порогуванню, що здобувач також визнає на сторінці 147.

Без аналізу чутливості важко відрізнити властивість даних від артефакту обраних порогів, хоча для скринінгового індексу, який за задумом переглядає фахівець, це обмеження не є критичним.

5. Набір даних для третього методу (п. 3.5.3) містить лише мітку депресивного чи недепресивного твіту і не містить ні ознак ПТСР, ні даних про реальну поведінку користувачів, тому, як вказано на сторінці 147, перевірено працездатність механізму агрегації, а не прогностичну здатність індексу. У наборах даних другого методу (COMSYS-T1) класу ПТСР також немає, тому «супутність» виявлюваних розладів саме ПТСР у роботі не перевірялась (с. 53, 147). Отже, для методу, у назві якого фігурує попередження соціально небезпечної поведінки громадян із ПТСР, експериментальне підтвердження обмежується якісною узгодженістю з міткою депресії та ручним переглядом дописів трьох профілів (рис. 4.17–4.19). Це коректно відображено у висновках, однак варто було б хоча б на невеликій вибірці, розміченій фахівцем, кількісно оцінити якість ранжування користувачів за індексом ризику.

6. Огляд методів пояснюваності у п. 1.6 (с. 47–52) є доволі широким, проте, як зазначено на сторінці 52, методи LIME, SHAP, Grad-CAM і контрфактичні пояснення в роботі не застосовуються і наведені лише для повноти огляду. Аналогічно модель із квантовим шаром (п. 3.2.2, 4.2.1) не входить до складу методу, досліджена в одному запуску, а різницю в 1 відсотковий пункт порівняно з моделлю без квантового шару здобувач сам не вважає підтвердженням ефекту (с. 112–113). Цей матеріал доцільно було б скоротити або винести в додаток, зосередивши виклад на складових, що формують наукову новизну.

Висловлені зауваження не знижують загальної позитивної оцінки дисертаційної роботи, оскільки не впливають на якість отриманих результатів, здебільшого стосуються обмежень, які здобувач сам виявив і зафіксував у п. 4.6, мають уточнювальний і рекомендаційний характер та можуть бути враховані здобувачем у подальших дослідженнях.

Загальний висновок.

Дисертаційна робота «Методи аналізу психічного стану в громадян із посттравматичним стресовим розладом за інтелектуальним аналізом користувацького контенту» є завершеною науково-дослідною працею і відповідає Стандарту вищої освіти України зі спеціальності 122 – «Комп'ютерні науки» для третього (освітньо-наукового) рівня вищої освіти, зокрема об'єктам вивчення та/або діяльності, пов'язаним із процесами збору, представлення, обробки, зберігання, передачі та доступу до інформації в комп'ютерних системах.

Дисертація присвячена розв'язанню актуальної науково-прикладної задачі аналізу психічного стану громадян із посттравматичним стресовим

розладом за інтелектуальним аналізом користувацького контенту. Розроблені в роботі методи дозволяють виявляти ознаки ПТСР у текстових повідомленнях користувачів з відмежуванням від тематично суміжних станів, ідентифікувати кілька психічних розладів одночасно без примусового вибору одного домінуючого стану та формувати пояснюваний часовий індекс ризику соціально небезпечної поведінки на рівні користувача. Це є важливим для підтримки фахівців і соціальних служб у ранньому виявленні осіб, які потребують допомоги, за умови дотримання етичних обмежень, сформульованих у роботі.

Оцінюючи дисертаційну роботу в цілому, вважаю, що за актуальністю теми, науковою новизною, обсягом виконаних досліджень, цінністю одержаних результатів і науково-теоретичним рівнем обґрунтованості результатів, робота цілком відповідає вимогам пунктів 6, 7, 8, 9 «Порядку присудження ступеня доктора філософії та припинення дії рішення разової спеціалізованої вченої ради закладу вищої освіти, наукової установи про присудження ступеня доктора філософії», затвердженого Постановою Кабінету Міністрів України від 12 січня 2022 р. № 44 (зі змінами, внесеними згідно з Постановами Кабінету Міністрів України № 341 від 21.03.2022, № 502 від 19.05.2023, № 507 від 03.05.2024, № 1240 від 07.10.2026), а її автор, здобувач Овчарук Олександр Миколайович, заслуговує на присудження ступеня доктора філософії за спеціальністю 122 – «Комп'ютерні науки» в галузі знань 12 – «Інформаційні технології».

Офіційний опонент
доктор фізико-математичних наук, професор,
професор кафедри математичної інформатики
факультету комп'ютерних наук та кібернетики
Київського національного університету
імені Тараса Шевченка _____ Олександр МАРЧЕНКО

Підпис проф. Марченко О.О.
засвідчую
Декан Кашпур О.Ф.

Київський національний університет імені Тараса Шевченка * Україна * Ідентифікаційний код 02070944
Факультет комп'ютерних наук та кібернетики