

Голові разової спеціалізованої
вченої ради PhD 13697
Хмельницького національного університету
доктору технічних наук, професору
Едуарду МАНЗІЮКУ

ВІДГУК

офіційного опонента на дисертаційну роботу
Віта Романа Віталійовича
за темою «Методи виявлення цифрової втоми у текстовому контенті
засобами штучного інтелекту»,
подану на здобуття наукового ступеня доктора філософії
з галузі знань 12 Інформаційні технології
за спеціальністю 122 Комп'ютерні науки

Актуальність теми дисертації.

Дисертаційна робота Віта Романа Віталійовича присвячена розробленню методів виявлення цифрової втоми у текстовому контенті засобами штучного інтелекту. Обрана тематика є актуальною, оскільки сучасні цифрові середовища стали невід'ємною складовою професійної, освітньої та соціальної взаємодії, а інтенсивне використання онлайн-комунікацій супроводжується зростанням когнітивного й емоційного навантаження на користувачів.

Наукова своєчасність роботи визначається міждисциплінарним характером поставленої задачі, що перебуває на перетині обробки природної мови, методів глибокого навчання, інтелектуального аналізу текстових даних та моніторингу психоемоційних станів користувачів цифрових середовищ. Важливим є те, що здобувач розглядає цифрову втому не лише як загальний наслідок тривалого перебування в цифровому середовищі, а як складний феномен, який потребує формалізованого виявлення за мовними ознаками у текстовому контенті. Суттєва складність досліджуваної задачі полягає в тому, що наявність у тексті тематики втоми, напруження чи негативних подій не завжди є прямим індикатором саме цифрової втоми як психоемоційного стану автора. Водночас латентні ознаки цифрового перевантаження можуть проявлятися без очевидних прямих згадок про цифровий контекст. Тому запропоноване в дисертації спрямування на розмежування тематичного змісту повідомлень і прихованих психологічних маркерів стану людини є методологічно виправданим та науково значущим.

Доцільним є також обраний здобувачем підхід до аналізу текстового контенту з урахуванням комунікативних сегментів, а не лише узагальнених

показників цифрової активності користувача. Така постановка задачі дає змогу перейти від поверхневого аналізу повідомлень до більш деталізованого виявлення джерел когнітивного навантаження, що відповідає сучасним вимогам до інтерпретованих інтелектуальних систем.

Соціальна значущість дослідження посилюється тим, що проблема цифрової втоми пов'язана з якістю освітнього процесу, продуктивністю праці, ментальним благополуччям користувачів і формуванням безпечного цифрового середовища. У цьому контексті дисертаційна робота узгоджується з актуальними науковими й суспільними пріоритетами, зокрема з питаннями цифрового добробуту, підтримки психічного здоров'я та відповідального використання цифрових технологій.

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 якості виявлення цифрової втоми у текстовому контенті з деталізацією за комунікативними сегментами засобами штучного інтелекту й подальша інтерпретація прийнятих рішень.

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

У *першому розділі* дисертації проведено аналіз методів, засобів та технологій для автоматизованого виявлення цифрової втоми та цифрового виснаження. Розглянуто сучасний стан предметної області тематичної класифікації текстової інформації для визначення комунікативних сегментів, проаналізовано ознаки цифрової втоми у текстовому контенті, а також підходи до візуальної аналітики та інтерпретації результатів роботи нейромережевих моделей. За результатами проведеного аналізу обґрунтовано доцільність застосування методів обробки природної мови та глибокого навчання для виявлення цифрової втоми у текстових даних, а також сформульовано мету і завдання дослідження.

У *другому розділі* розроблено формалізоване подання процесу виявлення цифрової втоми та цифрового виснаження засобами штучного інтелекту. Запропоновано концепцію роботи, у межах якої цифрова втома розглядається як сегментно зумовлений феномен, а цифрове виснаження – як системний стан, що формується внаслідок поширення проявів втоми між різними комунікативними сегментами цифрового профілю автора. Наведено модель виявлення цифрової втоми за текстовим контентом, охарактеризовано експериментальну базу і набори текстових даних, а також визначено засоби візуальної інтерпретації результатів на рівні комунікативних сегментів. Окремо слід відзначити розгляд критеріїв алгоритмічної прозорості та етичних обмежень моделювання психоемоційних станів, що є важливим для коректного використання результатів такого аналізу.

У *третьому розділі* наведено розроблені методи нейромережевого виявлення цифрової втоми та цифрового виснаження з візуальною інтерпретацією. Зокрема, запропоновано метод визначення комунікативних сегментів користувачів засобами штучного інтелекту, який забезпечує тематичну декомпозицію цифрового профілю автора та дає змогу розглядати окремі сегменти спілкування як самостійні джерела когнітивного навантаження. Розроблено метод виявлення цифрової втоми за комунікативними сегментами нейромережевими моделями глибокого навчання, що дозволяє перейти від глобального аналізу текстового масиву до локалізованого оцінювання проявів цифрової втоми в межах окремих тематичних контекстів. Також наведено метод виявлення цифрового виснаження з візуальною інтерпретацією за профілем автора, у якому локальні оцінки цифрової втоми узагальнюються в інтегральний профіль цифрового виснаження. Додатково розглянуто виявлення цільових об'єктів комунікативних сегментів, що забезпечує змістове пояснення отриманих результатів.

У *четвертому розділі* подано опис розробленого експериментального програмного забезпечення, яке реалізує запропоновані методи, та наведено результати їх експериментальної перевірки. Досліджено ефективність методу визначення комунікативних сегментів користувачів, методу виявлення цифрової втоми нейромережевими моделями глибокого навчання та методу виявлення цифрового виснаження за профілем автора. Показано, що запропонований підхід дозволяє формувати семантично відокремлені комунікативні сегменти, придатні для подальшого аналізу цифрової втоми, а використання трансформерних моделей забезпечує можливість автоматизованого виявлення її проявів у коротких текстових повідомленнях. Важливим результатом є підтвердження того, що цифрове виснаження не

зводиться до простого усереднення локальних оцінок або кількості негативно маркованих повідомлень, а потребує профільного, сегментно орієнтованого аналізу цифрової активності автора.

У висновках дисертації подано узагальнення отриманих результатів, зазначено виконані завдання, сформульовано основні наукові та практичні результати роботи, а також наведено відомості про практичне застосування розроблених методів. Список використаних джерел загалом відповідає предметній області дослідження та охоплює сучасні праці з обробки природної мови, тематичного моделювання, нейромережевого аналізу текстових даних і цифрової втоми. Додатки є доречними та доповнюють основний зміст роботи, зокрема в частині підтвердження публікацій, впровадження результатів, програмної реалізації та авторських свідоцтв.

Основні результати роботи:

1. Проведено аналіз методів, засобів і технологій для автоматизованого виявлення цифрової втоми та виснаження.

2. Розроблено метод визначення комунікативних сегментів користувачів засобами штучного інтелекту.

3. Розроблено метод виявлення цифрової втоми за комунікативними сегментами нейромережевими моделями глибокого навчання.

4. Розроблено метод виявлення цифрового виснаження за профілем автора з візуальною інтерпретацією прийнятих рішень.

5. Проведено експериментальне дослідження методів виявлення цифрової втоми та виснаження за комунікативними сегментами.

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, застосуванням адекватного методичного апарату, експериментальною перевіркою запропонованих методів, програмною реалізацією, 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 виявлення цифрової втоми та цифрового виснаження.

Наукова новизна отриманих результатів полягає в наступному:

1. Розроблено новий метод визначення комунікативних сегментів користувачів засобами штучного інтелекту, який на відміну від наявних підходів до аналізу цифрової активності, розглядає не глобальний цифровий профіль

автора, а множину локальних сегментів спілкування, що дає змогу інтерпретувати кожен сегмент як окреме джерело когнітивного навантаження.

2. Розроблено новий метод виявлення цифрової втоми за комунікативними сегментами нейромережевими моделями глибокого навчання, який на відміну від існуючих підходів, орієнтованих на глобальні показники активності користувача (загальний обсяг повідомлень, час онлайн, усереднений емоційний фон тощо), забезпечує посегментний аналіз цифрової втоми за текстовими даними, що дає змогу визначати стан цифрового виснаження як узагальнення неоднорідного впливу цифрової втоми за окремими комунікативними сегментами.

3. Розроблено новий метод виявлення цифрового виснаження з візуальною інтерпретацією за профілем автора, особливістю якого є використання результатів локального оцінювання цифрової втоми за комунікативними сегментами для побудови інтегрального профілю цифрового виснаження, що на відміну від існуючих підходів, орієнтованих переважно на агреговані показники цифрової активності, запропонований метод дає змогу розмежовувати локальні прояви цифрової втоми та критичний стан цифрового виснаження.

Теоретичне та практичне значення одержаних результатів.

Наукова цінність дисертаційної роботи полягає в тому, що здобувачем обґрунтовано, формалізовано та експериментально підтверджено підхід до виявлення цифрової втоми й цифрового виснаження у текстовому контенті засобами штучного інтелекту. На відміну від підходів, орієнтованих переважно на загальні показники цифрової активності, у роботі запропоновано посегментний аналіз цифрового профілю автора, що дозволяє розглядати окремі комунікативні сегменти як джерела когнітивного навантаження та відмежовувати локальні прояви цифрової втоми від критичного стану цифрового виснаження.

Практичне значення отриманих результатів полягає у програмній реалізації розроблених методів інтелектуального аналізу текстових даних. Розроблене програмне забезпечення забезпечує сегментування комунікативної активності користувача, виявлення цифрової втоми за комунікативними сегментами, формування інтегрального профілю цифрового виснаження автора та візуальну інтерпретацію отриманих результатів. Це підвищує не лише точність, а й інтерпретованість моніторингу стану користувача в цифровому середовищі.

Розроблені в роботі методи дозволили досягти поставленої мети дисертаційного дослідження, що полягала в підвищенні якості виявлення цифрової втоми у текстовому контенті з деталізацією за комунікативними сегментами та подальшою інтерпретацією прийнятих рішень.

Експериментальні результати підтверджують працездатність запропонованого підходу: для методу виявлення цифрової втоми за комунікативними сегментами отримано значення Precision 0.837, Recall 0.833 та F1-score 0.834.

Практична затребуваність результатів підтверджується їх упровадженням в освітній процес Воєнної академії імені Євгенія Березняка та Хмельницького національного університету, а також апробацією розробок у діяльності ГО «ІТ Кластер м. Хмельницького» і ПП «Авіві». 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 Віта Романа Віталійовича 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 Віта Р.В. 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. Дисертація написана з дотриманням принципів академічної доброчесності.

Оприлюднення результатів дисертаційної роботи.

За результатами досліджень дисертації опубліковано 12 наукових праць. Серед опублікованих наявні 4 статті у фахових наукових журналах України, включених на дату опублікування до переліку наукових фахових видань України категорії Б (4 з яких зараховано як 4 згідно із п.8 Постанови Кабінету Міністрів України від 12 січня 2022 р. № 44); 6 публікацій, які засвідчують апробацію матеріалів дисертації; 2 авторських свідоцтва. Тобто здобувач має 4 публікації, у яких викладені основні результати дисертації, чого достатньо згідно чинних вимог. У спільних публікаціях чітко зазначено особистий внесок дисертанта. Науковий рівень публікацій є високим. У всіх публікаціях здобувачем дотримано принципів академічної доброчесності.

Представлені публікації належним чином відображають результати проведених досліджень у межах дисертаційної роботи і не містять жодних порушень академічної чесності. Особистий внесок здобувача, що відображений у цих публікаціях і детально описаний у дисертації, підтверджує його авторство відповідних наукових висновків. Таким чином, наукові результати описані в дисертаційній роботі повністю висвітлені у наукових публікаціях здобувача.

Зауваження та побажання.

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

1. Проблема поняття «цифрова втома». Заявлено, що розробляються методи виявлення цифрової втоми та цифрового виснаження у текстовому контенті на рівні назви, мети, предмета, наукової новизни та висновків. Тобто робота подається саме як дослідження цифрової втоми. Але експериментальна база спирається не на корпуси, безпосередньо розмічені за цифровою втомою, а на суміжні психоемоційні стани: професійне виснаження, депресивні маркери, тривожність та прояви розладів.

На сторінці 58 прямо вказано, що для навчання й валідації моделей використовуються три набори даних: «Healthcare Workers' Burnout Tweets» для донавчання BERT-моделі з метою виявлення цифрової втоми, «Depression: Twitter Dataset + Feature Extraction» для валідації виявлення цифрової втоми за комунікативними сегментами та цифрового виснаження, а також «Students Anxiety and Depression Dataset» для валідації виявлення цільових об'єктів комунікативних сегментів.

У тексті роботи є кілька застережень, наприклад:

На сторінці 65 сказано, що клінічна депресія не є тотожною цифровій втомі, а DS2 не розглядається як джерело прямих міток цифрової втоми, а лише як корпус для перевірки переносності на суміжні психоемоційні стани,

На сторінці 68 це застереження ще сильніше: результати на DS2 треба трактувати як перевірку на суміжному стані, а не як безпосереднє підтвердження здатності моделі виявляти саме цифрову втому.

Є додаткове визнання, що психоемоційно несприятливо марковані повідомлення не є повним еквівалентом цифрового виснаження: на сторінці 155 вказано, що ці характеристики пов'язані, але не тотожні. Водночас нетотожність станів не перетворена на строгу експериментальну умову. Можна було б спробувати окремо довести, які мовні маркери є специфічними саме для цифрової втоми, а які лише відображають загальний дистрес.

Тобто питання «де тут саме втома?» так і не отримало відповіді. На підставі чого мітки burnout/depression/anxiety вважаються мітками цифрової втоми? Фактично, це одна велика проблема майже по всьому тексту роботи.

2. Як наслідок попереднього – використання DS2 як валідації цифрового виснаження. DS2 описаний як «Depression: Twitter Dataset + Feature Extraction». На сторінці 65 автор коректно зазначає, що клінічна депресія не є тотожною цифровій втомі, а корпус використовується як додатковий англomовний набір даних для перевірки переносності методів на суміжні психоемоційні стани. На сторінці 68 також сказано, що результати на цьому корпусі слід трактувати як

перевірку на суміжному стані, а не як безпосереднє підтвердження здатності моделі виявляти саме цифрову втому.

Однак у подальших розділах і висновках формулювання стають сильнішими без будь-яких підстав. На сторінці 166 зазначено, що ознаки цифрового виснаження виявлено у 25 із 52 авторів, тобто у 48,08 % вибірки. При цьому таблиця 4.8 і текст на сторінках 154–155 фактично зіставляють «профільне цифрове виснаження» з часткою повідомлень із проявами розладів.

Проблема полягає в тому, що автор одночасно визнає нетотожність депресивних/розладових міток і цифрової втоми, але у висновках формулює результат так, ніби цифрове виснаження було безпосередньо виявлене.

3. Пороги і агрегування результатів. У рукописі використано багаторівневу систему порогів. Але не видно ознак за якими вони були вибрані. Тобто є змістове обґрунтування бінаризації: автор пояснює, що локальний індекс має відображати не середню впевненість моделі, а частку повідомлень, для яких критерій втоми спрацював.

А) Тау (сторінка 53) взято як стандартне значення 0.5, але не показано, що саме це значення оптимальне для задачі.

Б) Альфа (стор 54) подано як значення 0.8, але не показано, як змінюються результати при інших значеннях. І на сторінці 142 у прикладі аналізу профілю використано порогові параметри і отримано Coverage = 46,7 %, але не показано, як змінився б висновок при іншому альфа.

В) Виглядає, наче гамма (сторінка 54), бета (сторінка 158) і ро-фет (сторінка 159) не мають чітко заданих числових значень.

Г) То як співвідносяться альфа-фет і ро-фет? На сторінці 159 ро-фет визначено як коефіцієнт відповідності навчального корпусу DS1 цільовому цифровому профілю D, але на сторінці 160 альфа-фет уже мінімально допустимий рівень відповідності. При цьому ро-фет має бути менше альфа-фет, тобто менше мінімуму. Це формули 4.4 і 4.5.

Д) Немає аналізу чутливості. Якщо змінити тау, гамму або альфу, кількість авторів, віднесених до групи «системного виснаження», може суттєво змінитися. Без аналізу чутливості важко відрізнити властивість даних від артефакту вибраних порогів.

4. Неоднозначний спосіб оцінювання. Coverage (сторінка 54) зважується за кількістю повідомлень. Це означає, що навіть один великий сегмент може непропорційно вплинути на інтегральну оцінку і створити враження системного поширення, хоча тематично проблема локалізована. І навпаки, кілька малих проблемних сегментів можуть бути недооцінені.

5. Підвищення якості. Заявлена у меті, і начебто підтверджена у висновках, але, наявність метрики predictive performance не означає доведення, що саме

запропонована сегментно-профільна схема підвищує якість порівняно з базою. У роботі не видно чіткого baseline.

6. У п. 4.3 наведено порівняння кількох трансформерних моделей, зокрема `vinai/bertweet-base`, `mental/mental-roberta-base`, `microsoft/deberta-v3-base` та `cardiffnlp/twitter-roberta-base`. При цьому модель `vinai/bertweet-base` визначено як найбільш збалансовану, хоча для `microsoft/deberta-v3-base` наведено вище значення F1-score – 0.843 проти 0.834. Було б доцільно формалізувати інтегральний критерій вибору базової моделі, оскільки `vinai/bertweet-base` має вищу Accuracy та кращу роздільну здатність за ROC-кривими, тоді як `microsoft/deberta-v3-base` демонструє вище значення F1-score.

7. У додатку В наведено програмні коди, що підтверджують практичну реалізацію окремих положень роботи. Водночас для повнішої відтворюваності експериментів доцільно було б подати розширені відомості про версії основних бібліотек.

Висловлені зауваження не знижують загальної позитивної оцінки дисертаційної роботи, оскільки не впливають на якість отриманих результатів, мають уточнювальний і рекомендаційний характер та можуть бути враховані здобувачем у подальших дослідженнях.

Загальний висновок.

Дисертаційна робота «Методи виявлення цифрової втоми у текстовому контенті засобами штучного інтелекту» 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 виявлення цифрової втоми та цифрового виснаження у текстовому контенті засобами штучного інтелекту. Розроблені в роботі методи дозволяють здійснювати інтелектуальний аналіз текстових повідомлень користувачів цифрових середовищ, визначати комунікативні сегменти, виявляти в них ознаки цифрової втоми та формувати інтегральний профіль цифрового виснаження автора з подальшою візуальною інтерпретацією отриманих результатів. Це є важливим для підвищення прозорості моніторингу стану користувача в цифровому середовищі, своєчасного виявлення ризиків цифрового перевантаження та запобігання переходу локальних проявів втоми у стан цифрового виснаження.

Оцінюючи дисертаційну роботу в цілому, вважаю, що за актуальністю теми, науковою новизною, обсягом виконаних досліджень, цінністю одержаних

