

УДК 004.8

Кашперук Т.Р., Мазурець О.В.

*Хмельницький національний університет*

## **МЕТОД ВИЗНАЧЕННЯ МОВНО-ДЕТЕРМІНОВАНОГО ІНДЕКСУ СУБ'ЄКТИВНОГО БЛАГОПОЛУЧЧЯ З ВИКОРИСТАННЯМ NLP**

*Запропоновано метод визначення мовно-детермінованого індексу суб'єктивного благополуччя з використанням NLP та відповідну програмну реалізацію. Метою інтелектуальної системи є визначення рівня задоволеності життям за описом користувача за допомогою методу SVM, який визначає оцінку. На підставі оцінки створюється опис стану користувача. Використання розробленого методу дає можливість користувачу отримати оцінку щодо свого стану та за потреби прийняти рішення щодо необхідності його покращення. Програма реалізація методу забезпечує виконання функцій: аналіз тексту та визначення оцінки задоволеності життям за описом; збір та обробка внесеної користувачем інформації; перегляд використаних датасетів; перегляд моделей та інформація щодо них; можливість внесення змін у датасети.*

*The method for determining a language-determined subjective well-being index using NLP and a corresponding software implementation are proposed. The purpose of the intelligent system is to determine the level of life satisfaction based on the user's description using the SVM method, which determines the assessment. Based on the assessment, a description of the user's state is created. Using the developed method allows the user to obtain an assessment of their state and, if necessary, make a decision on the need to improve it. The program implementing the method provides the following functions: text analysis and determination of the assessment of life satisfaction based on the description; collection and processing of information entered by the user; viewing used datasets; viewing models and information about them; the ability to make changes to datasets.*

Вже протягом тривалого часу люди усвідомлюють важливість розуміння свого емоційного стану та стеження за ним, оскільки емоційний стан безпосередньо впливає на психологічне благополуччя особи. Це усвідомлення призвело до створення окремої науки, яка вивчає та аналізує емоційний стан людини - психології. Наукові дослідження різних аспектів людської психіки та їх впливу на відчуття комфорту і задоволення є важливими елементами визначення психологічного комфорту. Центральними аспектами для визначення рівня задоволеності життям є стан і психологічний комфорт людини, а фахівці, які допомагають досягти цього комфорту, є психологами.

В деяких сферах діяльності, де використовуються алгоритмічні системи, вже відбувається активне впровадження ШІ, наприклад, у медичній сфері його використання полягає в автоматизації діагностики за зображеннями медичних знімків (наприклад, рентгенівських знімків), в сфері фінансів використовується для прогнозування ринкових трендів, управління портфелем інвестицій, в автомобільній промисловості - для розробки систем автопілоту та безпеки на дорозі. Психологія не є винятком.

ШІ вже широко використовується і продовжує активно вбудовуватись в психологічну практику. Наприклад, розробляються моделі, які аналізують сприйняття та обробку інформації людиною [1]. Як відзначає практикуючий психолог Десмонд Сох, ШІ допоможе автоматизувати рутинні завдання, такі як психометричні оцінки або IQ-тести, тим самим звільняючи час для інших видів діяльності [2]. Згідно з Робертом Солсо, у когнітивній психології проблематика ШІ розглядається переважно як інструмент самопізнання та рефлексії [3]. Крім того, ШІ можна використати для розпізнавання емоцій, що сприятиме визначенню рівня задоволеності життям.

Задоволеність життям є складним показником, що включає загальне щастя та задоволеність людини життєвими обставинами [4]. Рівень життєрадісності можна визначити за тестами, такими як, наприклад, ті, які розроблені англійським психологом В. Коулманом або за шкалою американського психолога Керола Дайана. Суть таких тестів полягає в аналізі їх результатів та формулювання висновків.

Отже, наведені дані підтверджують важливість розуміння та стеження за своїм емоційним станом для досягнення психологічного благополуччя. З розвитком інформаційних технологій відкриваються нові можливості в автоматизації рутинних завдань, таких як психометричні оцінки та розпізнавання емоцій, що визначить рівень задоволеності життям. Таким чином, автоматизація цих процесів сприятиме більш ефективному визначенню та підтримці психологічного комфорту і задоволеності життям.

Методи обробки природної мови (NLP) продемонстрували значний потенціал у задачах автоматизованого аналізу текстових даних [5, 6], що робить їх одним із ключових інструментів сучасних досліджень у сфері соціальних наук, психології та комп'ютерної лінгвістики [7, 8]. Значний обсяг комунікації людини – від приватних повідомлень до публічних дописів у соціальних мережах – представлений у текстовій формі [9], що перетворює текст на цінне джерело інформації про емоційний стан [10, 11], поведінкові патерни [12] та рівень суб'єктивного благополуччя особи.

Сучасні NLP-підходи ґрунтуються на використанні статистичних і нейромережових моделей, які здатні виявляти приховані семантичні та емоційні маркери, недоступні для класичних лінгвістичних методів [13, 14]. Моделі трансформерної архітектури, такі як BERT, RoBERTa, GPT та їхні україномовні адаптації, забезпечують контекстуально залежне кодування текстів, що дозволяє ефективно інтерпретувати багатовимірні сигнали: тональність висловлювань, риторичні конструкції, частоту лексем, індикатори емоційної валентності та когнітивного стилю [15]. Ці характеристики є релевантними для оцінювання таких психологічних категорій, як рівень життєвої задоволеності, суб'єктивний добробут чи емоційне благополуччя.

У контексті визначення мовно-детермінованого індексу суб'єктивного благополуччя NLP надає можливість об'єктивувати та формалізувати текстові ознаки, що відображають індивідуальні оцінки життєвої ситуації [16]. Застосування моделей класифікації, регресійних предикторів та Zero-/Few-Shot-підходів дозволяє автоматично зіставляти мовні патерни з відомими психологічними шкалами (такими як SWLS, PANAS, WHO-5), забезпечуючи відтворюваність і масштабованість оцінювання [17].

Перспективи використання NLP у цій сфері охоплюють:

- створення універсальних моделей оцінки благополуччя, здатних адаптуватися до різних соціокультурних контекстів;
- інтеграцію мультимодальних даних (текст + аудіо/відео) для підвищення точності визначення емоційного стану;
- розроблення індикаторів раннього виявлення психологічних ризиків, що можуть бути корисними в клінічній психології, HR-аналітиці, соціальних дослідженнях;
- використання генеративних моделей для побудови інтерпретованих пояснень, що покращує прозорість і валідність NLP-систем;
- автоматизацію аналізу великих обсягів соціальних даних, що відкриває можливість проводити масштабні дослідження суб'єктивного благополуччя на популяційному рівні.

Таким чином, NLP виступає ефективним інструментом не лише для описової інтерпретації текстів, а й для побудови кількісних моделей психологічних характеристик [18, 19]. Це робить можливим розроблення методів визначення мовно-детермінованих індексів суб'єктивного благополуччя, що забезпечують новий рівень глибини аналізу та відкривають перспективи для міждисциплінарних застосувань [20, 21].

Для більш чіткого розуміння підходу було спроектовано схеми та діаграму методу визначеності рівня задоволеності життям людини за текстовим описом.

Планування та відображення процесів змін кроків є одним з невід'ємних частин створення програми. Усі схеми та діаграми, отримані в результаті аналізу, слугуватимуть основою для подальшого розвитку програмного продукту. Ознайомившись з ними, виникає чітке уявлення про те, як саме функціонує дане програмне забезпечення. Наступна схема відображає метод, за допомогою якого буде визначатись рівень задоволеності життям.

На рисунку 1 вхідними даними методу визначення мовно-детермінованого індексу суб'єктивного благополуччя з використанням NLP є текстовий опис користувача та набори навчених класифікаторів, з яким у наступних етапах будуть здійснюватися операції.

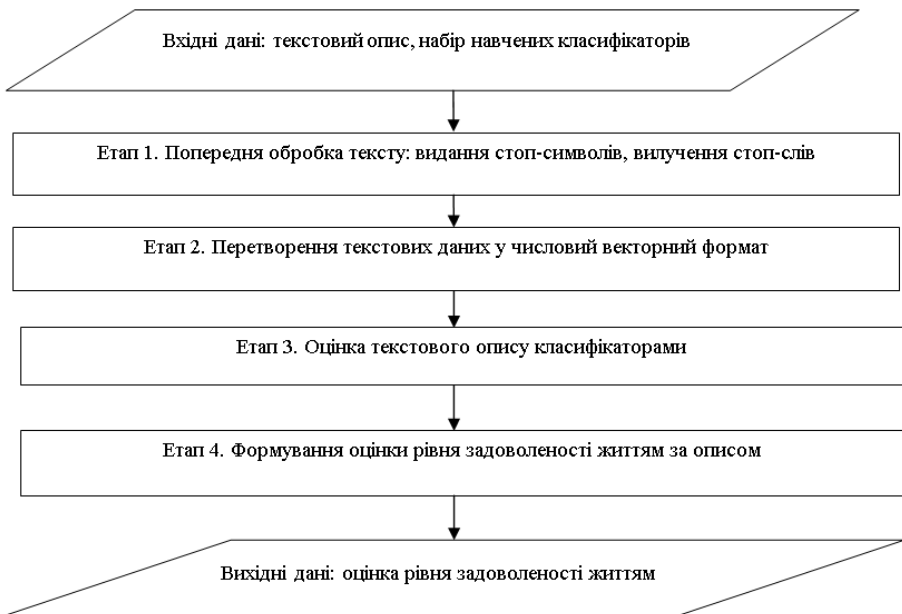


Рисунок 1 – Схема методу визначення мовно-детермінованого індексу суб'єктивного благополуччя

На першому етапі відображено підготовку даних, які включає в себе видалення стоп-символів та слів. Цей етап є обов'язковим перед використанням алгоритму SVM для аналізу текстових даних. Попередня обробка тексту дозволяє забезпечити зручні умови для подальшого використання алгоритму та оптимальну роботу класифікатора. Далі відбувається перетворення текстових даних у числовий векторний формат – векторизація даних, а саме: побудова векторного простору, де кожне слово представлене унікальним індексом. Третій крок – оцінка текстового

опису класифікаторами. Програма видає оцінку згідно навчених моделей та проаналізованого тексту. Останнім етапом є формування оцінки рівня задоволеності життям людини.

Вихідними даними є сама оцінка рівня задоволеності життям за шкалою від нуля до одиниці, в залежності від емоційної забарвленості опису. Якщо текст більше негативний, ніж позитивний – бал буде наближений до нуля, а якщо більш позитивний – до одиниці.

У схемі рисунку 2 показано більш деталізовано хід роботи алгоритму SVM: схема навчання класифікатора. Вхідними даними є датасет. Наступним кроком є вибір ядра. Вибір ядра був можливий з чотирьох основних типів ядра: лінійне, радіально базисне, поліноміальне та сигмоїдне. Для навчання було обрано лінійне ядро, оскільки воно має швидку обчислювальну складність та є ефективним для роздільних даних.

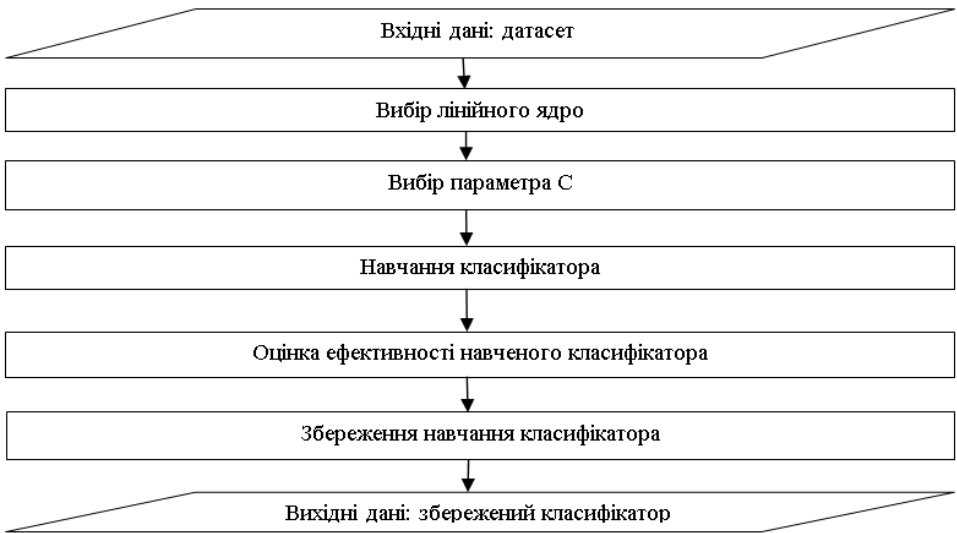


Рисунок 2 – Схема навчання типового класифікатора SVM

Наступним кроком є вибір параметра що визначає компроміс між рівнем точності на навчальному наборі та загальною здатністю моделі до узагальнення на нові дані. Значення параметра визначає рівень штрафу за помилки класифікації. Чим вище параметр, тим більший вплив має штраф на неправильну класифікацію, на процес підгонки моделі до навчальних даних. Після оптимізації параметрів, за допомогою вибраного ядра та оптимізованих параметрів SVM відбувається

побудова роздільної гіперплощини в просторі ознак, що найкращим чином розділяє класи даних.

Четвертий етап – навчання класифікатора. У випадку лінійної моделі навчання відбувається на навчальних даних шляхом максимізації ширини маржі між класами. Для цього використовується метод оптимізації, який мінімізує функцію втрат та враховує обмеження маржі.

П'ятий етап – це оцінка ефективності навченого класифікатора. Після завершення навчання модель оцінюється на тестових даних, щоб оцінити її точність та інші метрики ефективності, такі як точність, відновлення тощо. Останнім етапом є збереження навчання класифікатора. Вихідними даними є збережений класифікатор, який у подальшому буде використано у аналізі опису користувача для прорахунку оцінки рівня задоволеності життям.

За результатами досліджень було створено графіки та таблиці, що відображають результат функціонування програми, наведеної на рисунку 3.

Текст для аналізу:

У мене жакхливий настрій. Я постійно слухаю сумну музику. Боюсь, що так буде завжди. Зараз йде дощ...

Визначити рівень задоволеності життям

Результати аналізу:

Ймовірність радості: 0.02  
Ймовірність любові: 0.37  
Ймовірність страху: 1.00  
Ймовірність смутку: 1.00  
Коефіцієнт задоволеності рівня життя: 0.10  
Рівень задоволеності життям: Ви повністю незадоволені життям. Життя вам уже довгий час здається сірим, сумним та безнадійним, тому рекомендуємо звернутись до фахівця для отримання допомоги

Перегляд навчальних датасетів

Приклад моделей

Рисунок 3 – Інтелектуальна система визначення мовно-детермінованого індексу суб'єктивного благополуччя

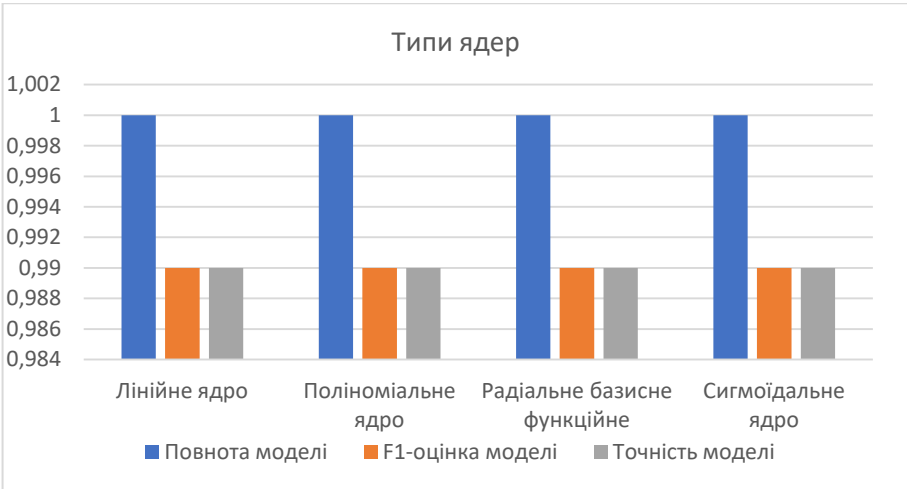


Рисунок 4 – Діаграма характеристик ядер

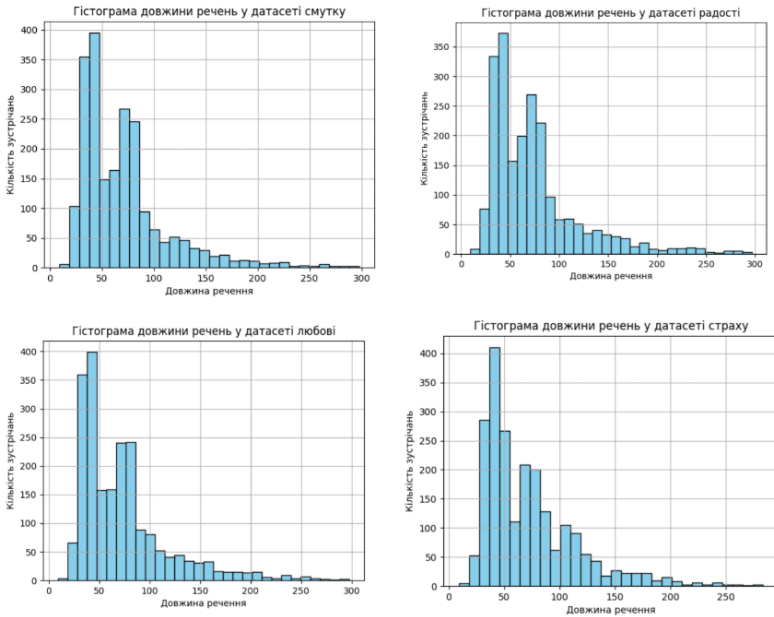


Рисунок 5 – Гістограми довжини речень датасетів

Для написання методу визначення рівня задоволеності життям людини було використано метод SVM, що має для навчання декілька типів ядра, і було протестовано різні варіанти ядра та їх вплив на кінцевий результат (рисунок 4).

З рисунка 4 видно, що характеристики у різних ядрах однакові, що свідчить що дані, що аналізуються, добре розділяються у просторі ознак за допомогою кожного з використовуваних ядер.

Також було проаналізовано довжину чотирьох датасетів, що зображено на рисунку 5

Усі чотири гістограми подібні одна до одного, що говорить про добре розподілення речень у датасетах та їх якісне навчання. На рисунку 6 зображено діаграми найпопулярніших слів у датасетах згідно проведеного аналізу.

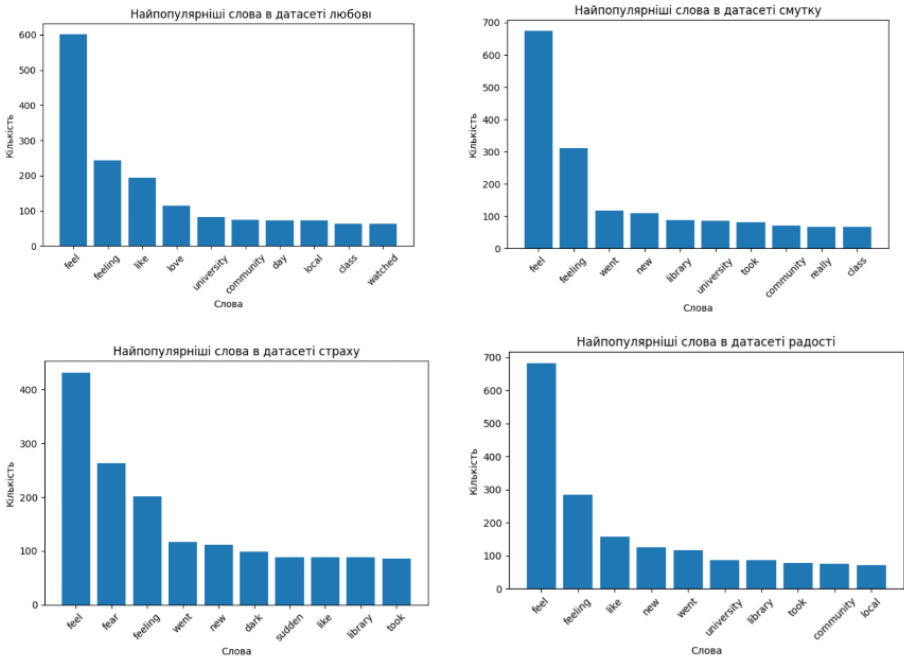


Рисунок 6 – Діаграми найпопулярніших слів у датасетах

Як видно з рисунка 6, кожна діаграма має свій набір слів. Далі було створено хмари слів на основі датасетів, які відображені на рисунку 7.



Рисунок 7 – Хмари слів

У кожній хмарі слів на рисунку 7 відображені від найбільш повторювані слова у датасетах до менш повторюваних, що видно з розмірах слів.

Отже, було спроектовано і реалізовано метод визначення мовно-детермінованого індексу суб'єктивного благополуччя з використанням NLP та відповідне ПЗ. Метою програми є визначення рівня задоволеності життям за описом користувача за допомогою методу SVM, який визначає оцінку. На підставі оцінки програма створює опис стану користувача. Використання розробленого ПЗ дає можливість користувачу отримати оцінку щодо свого стану та за потреби прийняти рішення щодо необхідності його покращення. Програма виконує наступні функції: аналіз тексту та визначення оцінки задоволеності життям за описом; збір та обробка внесеної користувачем інформації; перегляд використаних датасетів; перегляд моделей та інформація щодо них; можливість внесення змін у датасети.

Програма є перспективною, оскільки задоволеність життям та щастя людини прямо впливає на психологічний добробут і загальне здоров'я. Один із напрямків застосування може бути психологія. У майбутньому є шляхи удосконалення ПЗ, такі як збільшення обсягу даних для більш точної оцінки, збільшення кількості емоцій та інше.

## Перелік посилань

1. Psykholoh. Штучний інтелект. URL: <https://www.psykholoh.com/post/штучний-інтелект-це>
2. Liga.net. Психологія майбутнього. Як штучний інтелект допоможе у боротьби зі стресом та депресією. URL: <https://life.liga.net/poyasnennya/news/psihologiya-buduschego-kak-iskusstvennyy-intellekt-pomojet-v-borbe-so-stressom-i-depressiy>
3. Solso R. L. Cognitive Psychology: Pearson New International. United Kingdom: Pearson Education, 2013. 532 p.
4. Voxukraine. Віднайти щастя: визначення ключових факторів задоволеності життям серед українців. URL: <https://voxukraine.org/vidnajty-shhastya-vyznachennya-klyuchovyh-faktoriv-zadovolenosti-zhyttyam-sered-ukrayintsiv>
5. Денисенко Б.О., Молчанова М.О., Мазурець О.В. Інтелектуальна система виявлення дезінформації з застосуванням штучних нейронних мереж. Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024». 15-16 листопада 2024. Хмельницький, 2024. с. 167-174.
6. Шевчук П.О., Мазурець О.В. Метод аналізу достовірності текстових повідомлень для інтелектуальних систем кібербезпеки. Інформаційна, функційна і кібербезпека СКІФіК2024. Матеріали IV Всеукраїнської науково-технічної конференції. 29-30 листопада 2024. Харків. 2024. с. 118-119.
7. Mazurets O., Vit R. Practical Application of Method of Thematic Classification of Text Information Using LDA. Information Technology and Implementation (Satellite). Proceedings 11th International Conference. November 21, 2024. Kyiv, Ukraine. 2024. Pp. 151-152.
8. Віт Р.В., Мазурець О.В. Тематична класифікація текстової інформації засобами обробки природної мови. Збірник наукових праць XXIII Міжнародної наукової конференції «Нейромережні технології та їх застосування НМТіЗ-2024». 11-12 грудня 2024. Краматорськ-Тернопіль, ДДМА. 2024. с. 63-66.
9. 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.
10. Крак Ю.В., Дідур В.О., Молчанова М.О., Мазурець О.В., Собко О.В., Залуцька О.О., Бармак О.В. Метод виявлення політичної пропаганди в інтернет-контенті нейромережевими засобами обробки природної мови. Науковий журнал «Проблеми програмування». Київ, 2024, №2-3. с. 288-295.
11. Овчарук О.М., Мазурець О.В. Нейромережева архітектура з квантовим шаром для аналізу текстових повідомлень на прояви посттравматичного стресового розладу. Науковий журнал «Наука і техніка сьогодні». Київ, 2024. №13 (41). С. 1192-1204.
12. Юрченко Д.Ю., Овчарук О.М., Мазурець О.В., Шевчук П.О. Метод використання нейромережі гібридної архітектури для визначення емоційної тональності текстових повідомлень. Міжнародний науково-технічний журнал «Вимірювальна та обчислювальна техніка в технологічних процесах». № 2, 2025. с. 136-141.
13. Віт Р.В., Мазурець О.В. Метод виявлення психологічного цифрового перевантаження за аналізом текстових даних нейромережевими моделями глибокого навчання. Науковий

- журнал «Вісник Херсонського національного технічного університету». 2025. №2 (93). Т. 2. С. 107-114.
14. Мазурець О.В., Тимофієв І.А., Кліменко В.І., Тищенко О.О. Метод виявлення депресивного стану пов'язаного із навчанням у закладах освіти із використанням нейромережі дуальної архітектури. Науковий журнал «Вісник Херсонського національного технічного університету». 2024. №4 (91). С. 311-318.
15. Віг Р.В., Мазурець О.В. Підхід до тематичної класифікації текстової інформації засобами обробки природної мови. Науковий журнал «Наукові праці Донецького національного технічного університету», серія «Проблеми моделювання та автоматизації проектування». 2025. №1 (21). С. 94-99.
16. 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». № 1, 2025. с. 210-216.
17. Murava V., Zalutska O., Didur V., Mazurets O. Software architecture of information system for exchanging LLM thematic prompts. Global Trends in the Development of Information Technology and Science. Proceedings IV International Scientific and Practical Conference. June 25-27, 2025. Stockholm, Sweden. Pp. 121-127.
18. Віг Р.В., Мазурець О.В. Метод виявлення комунікаційних об'єктів як індикаторів цифрової втоми. Інтелектуальний метод виявлення цільових об'єктів предметної області для класифікації текстової інформації. Матеріали XIII Міжнародної науково-практичної конференції «Інформаційні управляючі системи та технології ІУСТ-ОДЕСА-2025». 24-26.09.2025. Одеса. 2025. С.119-121.
19. Залуцька О.О., Молчанова М.О., Мазурець О.В., Мельник О.І., Скрипник Т.К. Метод інтелектуального аналізу емоційної тональності текстової інформації для визначення поведінкових намірів нейромережевими засобами. Науковий журнал «Вісник Хмельницького національного університету» серія: Технічні науки. Хмельницький, 2023. №5 (325). Т.1. С. 67-73.
20. Мазурець О.В., Козенко О.В., Собко О.В. Метод автоматизованого підбору відповідей на користувачські запитання за семантичною подібністю. Матеріали XII Всеукраїнської науково-практичної конференції «Глушковські читання». Київ – 2023. с. 106-109.
21. Shevchuk P., Molchanova M., Mazurets O. Software for Text Messages Reliability Analysis Based on the Machine Learning Models Ensemble. Proceedings of IV International Scientific and Practical Conference «Innovative research and perspectives of the development of science and technology». January 29-31, 2024. Stockholm, Sweden. 2024. Pp. 347-354.