

УДК 004.8:004.9

DOI: 10.30837/0135-1710.2025.186.095

*С.Ф. ЧАЛИЙ, І.О. ЛЕЩИНСЬКА***МЕТОД ОЦІНКИ НЕГАТИВНИХ АСПЕКТІВ РІШЕННЯ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ В ЗАДАЧАХ ПОБУДОВИ ПОЯСНЕНЬ**

Об'єктом дослідження є процес оцінки рішення в інтелектуальній системі. Метою є розробка методу оцінки негативних аспектів рішень інтелектуальних систем з тим, щоб забезпечити формування збалансованих пояснень на основі інтеграції суттєвих негативних характеристик у ментальні моделі користувачів. Розроблено гібридний підхід до побудови ментальної моделі, який поєднує аналіз важливості та контекстуальний аналіз релевантності рішень з метою комплексної оцінки негативних властивостей вибраного користувачем рішення. Запропоновано метод оцінки негативних аспектів користувацького рішення при побудові ментальної моделі користувача, який інтегрує SHAP-аналіз важливості негативних властивостей з контекстним аналізом їхньої релевантності, що дає можливість автоматизованого виявлення та упорядкування негативних характеристик рішення для формування зрозумілих пояснень та підвищення довіри користувачів.

**1. Вступ**

Сучасні інтелектуальні системи (ІС) формують ефективні рішення для користувачів, використовуючи методи машинного навчання. Такі системи аналізують великі обсяги даних при формуванні персоналізованих рекомендацій у широкому спектрі предметних областей – від сфери електронної комерції й до області медичної діагностики. Проте наслідком використання технологій машинного навчання є виникнення проблеми непрозорості алгоритму прийняття рішень, коли користувачі ІС не мають інформації щодо логіки прийняття рішень інтелектуальною системою [1], [2].

Незрозумілість процесу формування рішення знижує довіру користувачів до ІС та обмежує їхню готовність використовувати рішення такої системи [3]. Це приводить до виникнення потреби у розробці методів пояснювального (зрозумілого) штучного інтелекту з тим, щоб забезпечити зрозумілість алгоритмічних рішень [4]-[6].

Одним із ключових підходів до вирішення проблеми непрозорості ІС у науковому напрямку пояснювального штучного інтелекту є побудова ментальних моделей користувачів [7]. Такі моделі відображають внутрішнє уявлення користувача ІС про процеси функціонування системи та її логіку прийняття рішень [8]. Однак традиційні підходи до побудови ментальних моделей орієнтовані переважно на представлення позитивних аспектів рішень ІС, не приділяючи достатньо уваги її негативним характеристикам [9]. Зокрема, підходи до інтерпретації важливості ознак вхідних даних не враховують аспект негативного сприйняття рішення користувачами [10]-[12]. Проте аналіз негативних аспектів рішень ІС має суттєве значення для формування збалансованого представлення про можливості та обмеження процесу формування рішення в такій системі. Користувачі ІС мають оцінити інформацію як про переваги запропонованого рішення, так і про його потенційні недоліки. Такий комплексний підхід дає можливість користувачеві підвищити довіру до інтелектуальної системи [13].

Таким чином, актуальною є проблема оцінки негативних аспектів рішень в ІС з тим, щоб забезпечити формування збалансованих пояснень в інтелектуальній системі на основі включення важливих негативних характеристик рішення до ментальної моделі користувача.

**2. Аналіз сучасних наукових публікацій і постановка проблеми дослідження**

Пояснювальний штучний інтелект орієнтований на розробку методів підтримки інтерпретації рішень ІС, що були отримані з використанням машинного навчання. Даний

напрямок започатковано програмою XAI DARPA [3]. Сучасні дослідження з пояснювального штучного інтелекту присвячені розробці методів як локальної, так і загальної інтерпретації рішень ІС [10]-[12], [14].

Представлений в [12] метод LIME призначений для локальної інтерпретації рішень ІС шляхом побудови спрощених моделей пояснень для конкретних прикладів рішень. Даний підхід виділяє ключові ознаки, що впливають на конкретне рішення інтелектуальної системи. Однак даний метод не враховує обмеження щодо глобальної узгодженості пояснень, а також не враховує контекст сприйняття рішень ІС користувачами [10].

Метод побудови пояснень SHAP (SHapley Additive exPlanations) використовує теорію кооперативних ігор і, на цій основі, забезпечує обґрунтований розподіл важливості ознак вхідних даних [10], [11]. Проте традиційний метод SHAP не враховує обмеження, пов'язані із негативним сприйняттям користувачами окремих аспектів рішення ІС.

Сучасні дослідження у сфері пояснювального штучного інтелекту орієнтовані на розробку методів, що враховують психологічні аспекти сприйняття пояснень користувачами. Останні представляються ментальними моделями [13], [14]. Ряд досліджень враховує також важливість адаптації пояснень до персональних особливостей користувачів ІС та контексту використання рішень [5], [15]-[17]. Класичні дослідження показали, що ментальні моделі відіграють ключову роль у взаємодії людини з технічними системами [18].

У контексті інтелектуальних систем ментальні моделі допомагають користувачам прогнозувати поведінку системи та приймати обґрунтовані рішення [8], [19]. Концептуальне представлення ментальної моделі щодо пояснення в ІС забезпечує структуровану репрезентацію знань про логіку прийняття рішень [8]. Ієрархічне, каузальне та функціонально-темпоральне представлення ментальних моделей в комплексі забезпечують можливість врахувати динаміку взаємодії користувача з ІС, наслідки альтернативних рішень [14], [20]. Принцип уточнення ментальних моделей на основі доповнення вхідних даних з урахуванням негативних характеристик рішення забезпечує підвищення релевантності пояснень [13], оскільки дозволяє врахувати позитивні, негативні та нейтральні аспекти, які впливають на вибір користувача [21], [22]. Виявлення негативних аспектів рішення дає можливість зрозуміти контекст незадоволення користувачів у сенсі обмежень щодо його використання і адаптувати пояснення [23].

Таким чином, існуючі підходи до побудови ментальних моделей та пояснень в ІС орієнтовані переважно на поясненні логіки рішення системи, не інтегруючи у пояснення важливих негативних характеристик цього рішення. Передумовою такої інтеграції є розробка методів виявлення та оцінки негативних характеристик рішень ІС на основі частково структурованих даних користувачів.

Зазначені обмеження існуючих методів свідчать про важливість розробки гібридного підходу, який би поєднував можливості існуючих методів ХАІ з аналізом негативних аспектів рішення для формування ментальних моделей користувачів ІС.

### **3. Мета і задачі дослідження**

Метою дослідження є розробка методу оцінки негативних аспектів рішень інтелектуальних систем з тим, щоб забезпечити формування збалансованих пояснень на основі інтеграції суттєвих негативних характеристик у ментальні моделі користувачів.

Для досягнення поставленої мети вирішуються такі задачі: розробити гібридний підхід до побудови ментальної моделі, який поєднує аналіз важливості та контекстуальний аналіз релевантності рішень з метою комплексної оцінки негативних властивостей вибраного користувачем рішення; розробити метод оцінки негативних аспектів користувацького рішення при побудові ментальної моделі користувача для автоматизованого виявлення та упорядкування негативних характеристик цього рішення.

#### 4. Гібридний підхід до побудови ментальної моделі з урахуванням негативних оцінок рішень інтелектуальної системи

Об'єктом дослідження є процес оцінки рішення в інтелектуальній системі. Предметом дослідження є принципи та методи оцінки негативних аспектів рішень інтелектуальних систем при формуванні користувацьких ментальних моделей для створення пояснень.

Запропонований підхід інтегрує результати аналізу важливості впливу властивостей вхідних даних з результатами контекстного аналізу релевантності цих властивостей для формування збалансованих ментальних моделей користувачів. Підхід ґрунтується на запропонованому в [20] принципі доповнення вхідних даних, згідно з яким доповнена  $j$ -та ментальна модель  $M_j$  представляється у вигляді множини позитивних (релевантних)

властивостей рішення  $V_j^+$  та множини негативних (нерелевантних) властивостей  $V_j^-$ , які обмежують використання рішення:  $M_j = V_j^+ \cup V_j^-$ .

Збалансована ментальна модель формує комплексне представлення користувача про рішення інтелектуальної системи, враховуючи як переваги, так і потенційні обмеження останнього. Таке рішення зменшує невизначеність при прийнятті рішень користувачем та, відповідно, створює умови для підвищення довіри до ІС.

Підхід реалізує послідовну оцінку негативних властивостей рішень ІС. На першому рівні виконується статистичний аналіз важливості характеристик даних для виявлення значущості останніх на основі їхнього внеску у формування загального сприйняття рішення. На другому рівні виконується контекстний аналіз релевантності рішення в сенсі оцінки суб'єктивної важливості для користувача негативних аспектів цього рішення для конкретних умов його використання. Процес формування ментальної моделі з урахуванням негативних аспектів рішення включає три основні етапи, представлені в табл. 1. Як вхідні дані при реалізації даного підходу використовується набір неструктурованих відгуків користувачів у текстовому форматі, дані щодо властивостей рішення ІС (предметна область, категорія, параметри, умови використання тощо). В рамках підходу як проміжні дані використовуються структуровані набори позитивних та негативних властивостей рішення ІС з кількісними оцінками частоти їх згадування у відгуках. На першому етапі розробленого підходу виконується підготовка даних для аналізу. На другому етапі формується комплексна оцінка важливості негативних аспектів на основі аналізу їхніх значущості та релевантності. На третьому етапі формується кінцева ментальна модель, що містить найсуттєвіші негативні характеристики рішення.

Представлений підхід забезпечує оцінку збалансованості ментальної моделі на основі співвідношення позитивних та негативних аспектів рішення ІС.

#### 5. Метод оцінки негативних аспектів рішення інтелектуальної системи

Розроблений метод базується на гібридному підході до побудови ментальної моделі.

Метод враховує негативні оцінки рішень інтелектуальної системи та включає такі етапи.

Етап 1. Попередня обробки неструктурованих вхідних даних. На даному етапі здійснюється стандартизація текстових даних, видалення шуму та виконання настроювання для виявлення емоційного забарвлення цих даних. Результатом етапу є множина структурованих даних, кожен елемент якої містить компоненти вхідного тексту та його настроювання-оцінку.

Етап 2. Класифікація властивостей рішення. На даному етапі виділяються підмножини позитивних  $V_j^+$  та негативних  $V_j^-$  властивостей рішення за результатами настроювання-аналізу на попередньому етапі.

Таблиця 1

## Основні етапи гібридного підходу

| Етапи                                              | Задача етапу                                                           | Реалізація етапу                                                                                                        | Результат етапу                                                 |
|----------------------------------------------------|------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------|
| 1. Структуризація вхідних даних                    | Перетворення неструктурованих відгуків користувачів у табличний формат | Сентимент-аналіз, класифікація негативних/позитивних аспектів                                                           | Множини структурованих властивостей вхідних даних $P^+$ , $P^-$ |
| 2. Гібридна оцінка негативних властивостей рішення | Оцінка значущості та релевантності негативних аспектів рішення         | – SHAP-аналіз значущості негативних аспектів рішення;<br>– контекстний аналіз релевантності негативних аспектів рішення | Ранжовані негативні аспекти з інтегрованими оцінками            |
| 3. Формування моделі                               | Інтеграція важливих негативних аспектів у ментальну модель             | Балансування ментальної моделі, відбір і валідація важливих негативних характеристик рішення                            | Доповнена ментальна модель                                      |

Етап 3. SHAP-аналіз важливості негативних властивостей. На даному етапі на основі теорії кооперативних ігор формується оцінка внеску кожної негативної характеристики у формування загального сприйняття рішення користувачами. Використовується метод SHAP. Оцінка кожної із властивостей визначається через обчислення її SHAP-значення [10] та подальшу нормалізацію цього значення. Результатом етапу є сума  $S_j^-$  SHAP-значень  $s_{i,j}^-$  всіх негативних властивостей:  $S_j^- = \sum_i s_{i,j}^-$ .

Етап 4. Контекстний аналіз негативних властивостей. На даному етапі враховується частота згадування  $f_{i,j}^-$  кожної негативної властивості у відгуках користувачів, оцінка впливу властивості на можливість використання рішення  $p_{i,j}^-$ , що отримана за результатами сентимент-аналізу, а також оцінка впливу на загальний користувацький досвід  $u_{i,j}^-$ . При оцінці  $p_{i,j}^-$  враховується вплив на функціональність (чи буде працювати рішення без відповідної функції), а також частота згадування про проблему у вхідних даних. При оцінці  $u_{i,j}^-$  враховується частота виникнення проблеми, а також її вплив на досвід використання рішення. Оцінка  $w_{i,j}^-$  для кожної властивості нормалізується зі з використанням змінної  $n_{i,j}^- = \max_i (f_{i,j}^- p_{i,j}^- u_{i,j}^-)$ :

$$w_{i,j}^- = \frac{f_{i,j}^- p_{i,j}^- u_{i,j}^-}{n_{i,j}^-}. \quad (1)$$

Результатом етапу є сумарна контекстна оцінка  $W_j^- = \sum_j w_{i,j}^-$ .

Етап 5. Формування комбінованої оцінки рішення ІС. На даному етапі визначається співвідношення між SHAP-оцінкою  $S_j^-$  та контекстною оцінкою  $W_j^-$ . Дане співвідношення порівнюється з пороговим значенням  $E$ :

$$\frac{S_j^-}{W_j^-} > E. \quad (2)$$

Такий підхід дає можливість оцінити вплив найважливіших негативних властивостей рішення ІС з урахуванням контексту їх використання. Тобто дане співвідношення дає можливість оцінити баланс представлення рішення з точки зору користувача.

Етап 6. Відбір ключових негативних аспектів рішення. На даному етапі формується підмножина рішень  $V_j^{--}$ , для яких співвідношення  $\frac{s_{i,j}^-}{w_{i,j}^-}$  менше, ніж значення (2), тобто відбираються найсуттєвіші негативні властивості:

$$V_j^{--} = \left\{ v_{i,j}^- : \frac{S_j^-}{W_j^-} > \frac{s_{i,j}^-}{w_{i,j}^-} \right\}. \quad (3)$$

Такий підхід дає можливість сфокусувати увагу лише на ключових обмеженнях рішення ІС та уникнути інформаційного перевантаження користувача і тим самим підвищити довіру до такого рішення.

Етап 7. Формування доповненої ментальної моделі рішення. На даному етапі ментальна модель  $j$ -го рішення представляється через множини позитивних та ключових негативних властивостей, що створює збалансоване представлення користувача про рішення ІС:

$$M_j = V_j^+ \cup V_j^{--}. \quad (4)$$

Отримана ментальна модель відображає комплексне представлення рішення, що включає переваги та суттєві обмеження даного рішення. Відповідно, така модель створює умови щодо обґрунтованого вибору рішення користувачем.

#### 6. Експериментальна перевірка розробленого методу

Експериментальну перевірку методу виконано з використанням користувацьких відгуків 2025 року про мобільний телефон на платформі систем електронної комерції. Експериментальна вибірка складалася із 65 відгуків. З них 45 відгуків містили інформацію про негативні аспекти даного телефону. Файл з відгуками містив ID користувача, дату публікації, оцінку телефону за п'ятибальною шкалою, текст відгуку, переваги та недоліки продукту, виділені користувачем властивості продукту.

Результати етапу 1 методу було представлено показниками негативних властивостей рішення. У відгуках було виділено 15 типів негативних аспектів, зокрема швидкість роботи, автономність батареї, якість камери, нагрів пристрою. Кожна негативна властивість отримала числовий показник відповідно до частоти згадування у множині відгуків та оцінки його важливості в контексті досвіду користувача. Діапазон оцінок властивостей визначався після нормалізації цих показників. При проведенні експерименту він становив 0,227–0,06. Показник з найвищим значенням характеризує властивість з найбільшою частотою негативних відгуків, що має найсуттєвіший вплив на загальну оцінку користувача. Показник з мінімальним значенням відповідає негативним властивостям, що мають найменше відгуків і, отже, мають незначний вплив на формування враження користувачів від продукту. На третьому етапі було отримано значення  $S_j^- = 0,9$  сумарного показника негативних властивостей. Таке високе SHAP-значення свідчить про суттєвий внесок негативних властивостей у загальну оцінку. На етапі 4 отримано значення

контекстної оцінки  $W_j^- = 0,346$ . Дане значення враховувало ступінь впливу на користувацький досвід, частоту виникнення проблеми та рівень впливу на рішення про придбання цієї моделі. Зокрема, оцінка швидкості роботи системи становила 0,12 внаслідок її суттєвого впливу на повсякденне використання телефону. Оцінка нагріву телефону під навантаженням становила 0,05 внаслідок незначного впливу на повсякденне використання пристрою. Найсуттєвішими властивостями рішення, що мали найбільший негативний вплив, були швидкість роботи системи, час автономної роботи батареї, якість зйомки камерою при слабкому освітленні, нагрів телефону під навантаженням,

співвідношення ціна-якість. Співвідношення  $\frac{S_j^-}{W_j^-} = 2,6$  свідчить, що розроблений метод

сепарує найкритичніші негативні властивості, які мають найбільший вплив на досвід користувача, від всієї сукупності негативних факторів.

### 7. Обговорення результатів

Розроблений метод оцінки негативних аспектів рішення інтелектуальної системи при побудові користувацької ментальної моделі відрізняється від існуючих комбінуванням оцінок сумарного та контекстно-орієнтованого впливу негативних аспектів рішення на сприйняття цього рішення користувачем системи, що дає можливість побудувати в рамках ментальної моделі комплексне представлення рішення, що включає переваги та суттєві обмеження при його використанні. Така інтеграція властивостей підвищує релевантність пояснень в інтелектуальних системах, що забезпечує можливість обґрунтованого вибору рішення користувачем.

Порівняння розробленого методу з методом оцінки впливу значень вхідних даних LIME, який орієнтований на пояснення впливу властивостей вхідних даних на рішення ІС, дало можливість зробити висновок про принципову відмінність між даними підходами. Метод LIME формує локальне пояснення для конкретного прикладу, щоб визначити вплив вхідних даних. Розроблений метод враховує вплив негативних характеристик на всьому наборі вхідних даних, тобто узагальнює особливості використання рішень багатьма користувачами. Тобто LIME оцінює алгоритм роботи моделі, а запропонований метод оцінює властивості рішення. Аналогічно, метод SHAP також пояснює причини, чому модель прийняла відповідне рішення і не дає можливість врахувати контекст використання рішення. На відміну від методу SHAP, який виконується на етапі 3 і дає можливість узагальнити вплив вхідних даних на рішення, розроблений метод оцінює результати SHAP у контексті використання рішення ІС.

Отримане при проведенні експерименту співвідношення  $\frac{S_j^-}{W_j^-} = 2,6$  підтверджує

ефективність виділення ключових проблемних властивостей рішення. Тобто близько чверті від усіх негативних властивостей має найбільший вплив на сприйняття рішення користувачем. Таке значення співвідношення свідчить, що лише незначна кількість недоліків призводить до значної кількості скарг користувачів.

Обмеження на використання даного методу пов'язані із релевантністю і повнотою вхідних даних, оскільки відгуки можуть бути сформовані в результаті шилінг-атак. Останні направлені на примусову зміну рейтингів конкурентних рішень. неструктурованих вхідних даних

Напрямки подальших досліджень щодо оцінки негативного впливу характеристик рішення пов'язані із розробкою моделей оцінки загального та контекстно-орієнтованого

впливу негативних властивостей з використанням машинного навчання, що дасть можливість використовувати інші структури вхідних даних, в тому числі мультимодальні дані.

## 8. Висновки

Запропоновано гібридний підхід до побудови ментальної моделі на основі оцінки негативних властивостей рішення інтелектуальної системи, який використовує комбіновану оцінку негативних властивостей вхідних даних для відбору цих властивостей при побудові ментальної моделі рішення. Такий підхід забезпечує прийняття користувачем обґрунтованого рішення з урахуванням його ключових негативних властивостей.

Запропоновано метод оцінки негативних аспектів рішення інтелектуальної системи при побудові ментальної моделі користувача, який містить етапи попередньої обробки вхідних неструктурованих даних, SHAP-аналізу важливості негативних властивостей вхідних даних, контекстного аналізу негативних властивостей, формування комбінованої оцінки рішення ІС. Метод дає можливість виявляти та упорядковувати негативні властивості рішення інтелектуальної системи, що створює умови для побудови зрозумілих для користувача пояснень та підвищення довіри користувачів внаслідок представлення потенційних обмежень для рішень ІС.

## Перелік посилань:

1. Miller T. Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence. 2019. Vol. 267. P. 1-38.
2. Salih A.M., Amin R., Shanta S.S. A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME. arXiv preprint arXiv:2305.02012. 2023.
3. Gunning D., Aha D.W. DARPA's explainable artificial intelligence (XAI) program. AI Magazine. 2019. Vol. 40. № 2. P. 44-58.
4. Sun Q., Liu Y., Chen H. et al. Explainable Artificial Intelligence for Medical Applications: A Review. arXiv preprint arXiv:2412.01829. 2024.
5. Cheung J.C., Wong A., Liu H. The effectiveness of explainable AI on human factors in trust and decision making. Nature Scientific Reports. 2025. Vol. 15. Art. 4189.
6. Sheridan H., Clarke D., Murphy C. Exploring Mental Models for Explainable Artificial Intelligence. HCHI 2023 Conference Proceedings. 2023.
7. Johnson-Laird P.N. Mental models and human reasoning. Proceedings of the National Academy of Sciences. 2013. Vol. 110. № 45. P. 18243-18250.
8. Чалий С.Ф., Лещинська І.О. Концептуальна ментальна модель пояснення в системі штучного інтелекту. Бюлетень Національного технічного університету "ХПІ". Серія: Системний аналіз, управління та інформаційні технології. 2023. № 1. С. 41-46.
9. Doshi-Velez F., Kim B. Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608. 2017.
10. Salih A.M., Amin R., Shanta S.S. A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME. arXiv preprint arXiv:2305.02012. 2023.
11. Lundberg S.M., Lee S.I. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017. P. 4768-4777.
12. 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. 2016. P. 1135-1144.
13. Чалий С.Ф., Лещинська І.О. Уточнення ментальної моделі рішення користувача інтелектуальної системи шляхом доповнення вхідних даних. Автоматизовані системи управління та прилади автоматики. 2024. № 188. С. 4-12.
14. Чалий С.Ф., Лещинський В.О., Лещинська І.О. Каузальна ментальна модель для пояснення в інтелектуальних системах. Автоматизовані системи управління та прилади автоматики. 2024. № 189. С. 12-20.
15. Sun Q., Liu Y., Chen H. et al. Explainable Artificial Intelligence for Medical Applications: A Review. arXiv preprint arXiv:2412.01829. 2024.
16. Jeong S., Miller T., Park J. AI Mental Models & Trust: The Promises and Perils of User Mental Models. EPIC Conference Proceedings. 2024. P. 234-251.
17. Dangol A., Peters C., Sharma K. Children's Mental Models of AI Reasoning. arXiv preprint

arXiv:2505.16031. 2025.

18. Norman D.A. Some observations on mental models. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 7–14). Lawrence Erlbaum Associates. 1983.

19. Чалий С.Ф., Лещинська І.О. Функціонально-темпоральне представлення ментальної моделі користувача для генерації пояснень. *Автоматизовані системи управління та прилади автоматики*. 2025. № 190. С. 15-24.

20. Чалий С.Ф., Лещинська І.О. Принципи побудови ментальних моделей зовнішніх користувачів інтелектуальних систем. *Вісник Національного технічного університету "ХПІ"*. 2024. № 7. С. 89-95.

21. Liu B. Sentiment analysis and opinion mining. *Synthesis lectures on human language technologies*. 2012. Vol. 5. № 1. P. 1-167.

22. Ahmed M.T., Rhaman S., Kader A. A systematic review of explainable artificial intelligence for spectroscopic models. *Computers and Electronics in Agriculture*. 2025. Vol. 228. Art. 109462.

23. Ullah N., Khan J.A., El-Sappagh S. A novel explainable AI framework for medical image analysis. *Medical Image Analysis*. 2025. Vol. 97. Art. 103241.

Надійшла до редколегії 18.09.2025 р.

**Чалий Сергій Федорович**, доктор технічних наук, професор, професор кафедри ІУС ХНУРЕ, м. Харків, Україна, e-mail: [serhii.chalyi@nure.ua](mailto:serhii.chalyi@nure.ua); ORCID: <https://orcid.org/0000-0002-9982-9091>

**Лещинська Ірина Олександрівна**, кандидат технічних наук, доцент, доцент кафедри ПІ ХНУРЕ, м. Харків, Україна, e-mail: [iryna.leshchynska@nure.ua](mailto:iryna.leshchynska@nure.ua), ORCID: <https://orcid.org/0000-0002-8737-4595>