УДК 004.891.3 **DOI: 10.20998/2079-0023.2022.02.15**

С. Ф. ЧАЛИЙ, В. О. ЛЕЩИНСЬКИЙ

МОДЕЛЬ ПОЯСНЕННЯ В ІНТЕЛЕКТУАЛЬНІЙ СИСТЕМІ НА ЛОКАЛЬНОМУ, ГРУПОВОМУ ТА ГЛОБАЛЬНОМУ РІВНЯХ ДЕТАЛІЗАЦІЇ

Предметом дослідження є процеси формування пояснень в інтелектуальних інформаційних системах. В сучасних інтелектуальних системах використовуються методи машинного навчання. Процес отримання рішення, сформованих на основі таких методів, є зазвичай непрозорим для користувача. Внаслідок такої непрозорості користувач може не довіряти тим рішенням, які запропонувала інтелектуальна система. Це знижує ефективність її використання. Для підвищення прозорості рішень використовуються пояснення. Пояснення представляється знаннями щодо причин формування результату в інтелектуальній системі, а також щодо причин окремих дій у процесі формування результату. Також пояснення може містити знання щодо впливу окремих функцій на отриманих інтелектуальною системою результат. Тому пояснення доцільно формувати на різних рівнях деталізації з тим, щоб показати як узагальнені причини та впливи на отримане рішення, так і причини вибору окремих проміжних дій. Мета роботи полягає в розробці узагальненої моделі пояснення з урахуванням станів та рівнів деталізації процесу прийняття рішення в інтелектуальній системі для побудови пояснень на основі відомих даних щодо послідовності станів та властивостей цих станів. Для досягнення мети вирішуються такі задачі: структуризація властивостей пояснень; визначення можливостей підходів до побудови пояснень на основі станів та структури процесу формування рішення, а також на основі вхідних даних; побудова моделі пояснення. Висновки. Запропоновано узагальнену модель пояснення в інтелектуальній системі для локального, групового та глобального рівнів деталізації процесу прийняття рішення .Модель представляється упорядкованою послідовністю зважених залежностей між подіями або станами процесу прийняття рішення. Модель орієнтовано на представлення можливості в рамках глобального пояснення виділити локальне пояснення та представити ланцюжок групових пояснень між подіями отримання вхідних даних та результуючого рішення. У практичному плані запропонована модель призначена для побудови пояснень з використанням підходів на основі спрощення процесу функціонування інтелектуальної системи та на основі виділення впливу окремих функцій та дій на кінцевий результат. Додаткові можливості моделі пов'язані із деталізацією подій процесу прийняття рішення з виділення окремих змінних, які характеризують стан цього процесу, що дає можливість формувати пояснення на основі використання відомих концепцій та понять у предметній області.

Ключові слова пояснення; інтелектуальна інформаційна система; залежності; рівні деталізації пояснень причинно-наслідкові зв'язки.

S. CHALYI, V. LESHCHYNSKYI

AN EXPLANATION MODEL IN AN INTELLIGENT SYSTEM AT THE LOCAL, GROUP AND GLOBAL LEVELS OF DETAIL

The subject of research is the process of formation of explanations in intellectual information systems. Machine learning methods are used in modern intelligent systems. The process of obtaining the solution formed on the basis of such methods is usually opaque to the user. As a result of such opacity, the user may not trust the solutions proposed by the intelligent system. This reduces the efficiency of its use. Explanations are used to increase the transparency of decisions. The explanation is represented by knowledge about the reasons for the formation of the result in the intellectual system, as well as about the reasons for individual actions in the process of formation of the result. Also, the explanation may contain knowledge about the influence of individual functions on the results obtained by the intelligent system. Therefore, it is advisable to form an explanation at different levels of detail in order to show both the generalized reasons and effects on the obtained decision, as well as the reasons for choosing individual intermediate actions. The purpose of the work is to develop a generalized model of explanation considering the states of the decision-making process in an intelligent system to build explanations based on known data regarding the sequence of states and the properties of these states. To achieve the goal, the following tasks are solved: structuring the properties of explanations; determining the possibilities of approaches to building explanations based on the states and structure of the decision-making process, as well as on the basis of input data; construction of an explanatory model. Conclusions. A generalized model of explanation in an intelligent system for local, group and global levels of detail of the decision-making process is proposed. The model is represented by an ordered sequence of weighted dependencies between events or states of the decision-making process. The model is focused on presenting the possibility to highlight a local explanation within the framework of a global explanation and to present a chain of group explanations between the events of obtaining input data and the resulting decision. In practical terms, the proposed model is intended for the construction of explanations using approaches based on the simplification of the process of functioning of the intelligent system and on the basis of highlighting the influence of individual functions and actions on the final result. Additional capabilities of the model are related to the detailing of the events of the decision-making process from the selection of individual variables that characterize the state of this process, which makes it possible to form an explanation based on the use of known concepts and concepts in the subject area.

Keywords: explanation; intelligent information system; dependencies; levels of detail of explanations cause-and-effect relationships.

Вступ. При побудові сучасних інтелектуальних інформаційних систем широко застосовуються методи машинного навчання з використанням об'ємних наборів даних [1]. В результаті моделі, які лежать в основі функціонування таких інтелектуальних систем, стають дуже складними для розуміння і непрозорими для користувача. Відповідно, користувач не завжди може довіряти рішенням, які запропонувала така інтелектуальна система [2].

Тому користувачі інтелектуальних систем потребують пояснення щодо отриманих рішень, щоб довіряти цим рішенням та ефективно їх використовувати [3]. У науковому аспекті пояснення дають можливість виявити упередженість у отриманих моделях. Така упередженість пов'язана із використанням для навчання наборів даних, в яких зафіксовано упередженість у рішеннях людей. Наприклад, проблема упередженості за певними ознаками може виникати при прийомі на роботу з використанням інтелектуальної системи підтримки прийняття рішень [4].

У практичному аспекті пояснення дають можливість оцінити ризики впровадження інтелектуальних

інформаційних технологій, обґрунтовуючи як відповідні послідовність дій при реалізації останніх, так і отримані результати. Слід також зазначити, що користувачі мають отримувати пояснення у різній формі, в залежності від задачі, що вирішується, а також від їх рівня підготовки та знань [5].

Зазначене свідчить про важливість задачі інтерпретації для користувача процесу формування рішення в інтелектуальній системі. Якщо алгоритми та моделі, що лежать в основі роботи системи, є непрозорими, то вони мають бути доповнені поясненнями.

Аналіз останніх досліджень і публікацій.

Дослідження щодо можливостей побудови пояснень в сучасних інтелектуальних системах були інтегровані та значно інтенсифіковані в рамках програми XAI [3]. Дослідження проводяться в напрямках визначення когнітивних аспектів пояснень [6], побудови інтелектуальних систем, функціонування яких може бути інтерпретовано безпосередньо [7, 8], формування пояснень з використанням каузальних та темпоральних залежностей [9–11].

Однак існуючі підходи не приділяють достатньо уваги побудові пояснень з урахуванням рівня їх деталізації. В той же час, виділення пояснень на відповідних рівня деталізації пояснень є необхідною умовою для їх представлення різним категоріям користувачів – розробникам, аудиторам, експертам в предметній області, кінцевим користувачам. Зазначене свідчить про важливість побудови узагальненої моделі пояснення з урахуванням рівнів його деталізації з тим, щоб адаптувати пояснення до потреб різних категорій користувачів.

Мета та задачі дослідження. Мета статті полягає у розробці узагальненої моделі пояснення з урахуванням станів та рівнів деталізації процесу прийняття рішення в інтелектуальній системі для побудови пояснень на основі відомих даних щодо послідовності станів та властивостей цих станів.

Для досягнення мети вирішуються такі задачі:

– структуризація властивостей пояснень;

– визначення можливостей підходів до побудови пояснень на основі станів та структури процесу формування рішення, а також на основі вхідних даних;

– побудова моделі пояснення.

Властивості пояснень в інтелектуальних системах. Інтелектуальні системи з точки зору зрозумілості їх роботи підрозділяються прозорі та непрозорі.

Перші характеризуються тим, що процес прийняття рішення є зрозумілим для користувача. Тобто людина може зрозуміти дії системи, реалізовані функції без потреби у додаткових поясненнях.

Зрозумілістю системи є її властивість представляти знання, на основі яких вона отримує рішення, у зрозумілій для користувача формі.

З урахуванням характеристик зрозумілості та незрозумілості інтелектуальні системи поділяються на системи з можливістю інтерпретації та системи з можливістю пояснення [7–11].

Системи першого типу використовують моделі, які описують процес прийняття рішень у зрозумілій для користувача формі.

Інтелектуальні інформаційні системи з можливістю пояснення містять додатковий інтерфейс для людини, який призначений для інтерпретації отриманих в інтелектуальній системі рішень та безпосередньо процесу прийняття рішень. Загальну схему, що відображає зв'язок наведених властивостей, представлено на рис. 1.

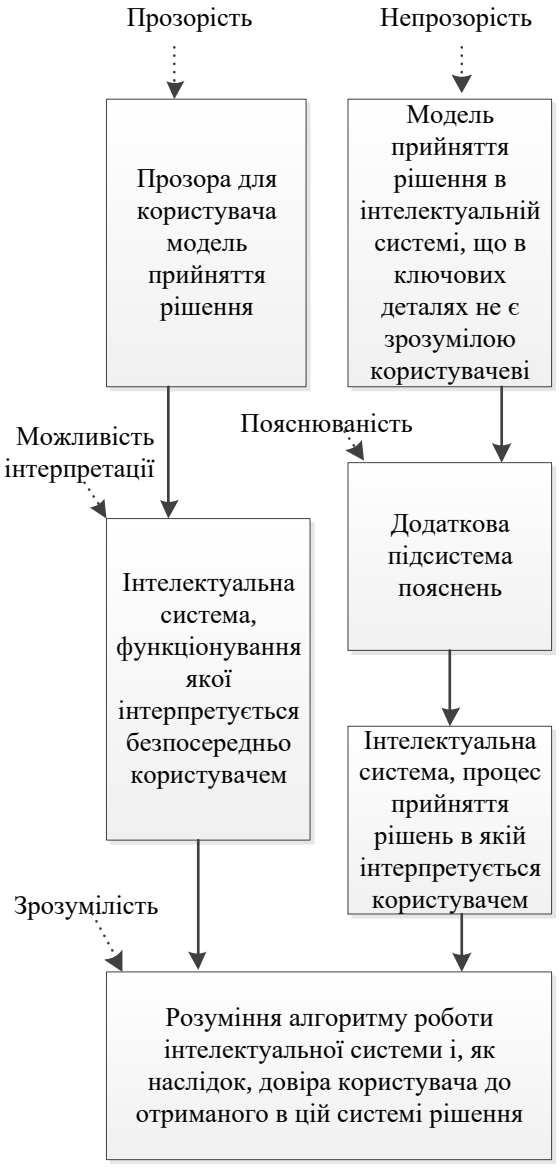

Рис. 1. Зв'язок властивостей інтерпретованості та пояснюваності інтелектуальної системи

Побудова пояснень в інтелектуальній системі відбувається в рамках двох ключових напрямків

– пояснення на основі станів та структури інтелектуальної системи;

– пояснення з використанням вхідних даних.

В рамках першого напрямку виділяються такі підходи:

– спрощення процесу роботи системи;

– визначення впливу окремих функцій на результат роботи інтелектуальної системи;

– використання відомих концепцій та понять у предметній області.

Ключові відмінності даних підходів наведено у табл. 1.

Підхід до побудови пояснень на основі спрощення знань щодо процесу роботи інтелектуальної системи використовує узагальнені, експериментально підтверджені закономірності, які відображають процеси та явища у предметній області але не використовують модель предметної області. Наприклад, закономірності виду «люди частіше хворіють на грип восени», без урахування природи таких захворювань.

Таблиця 1 – Пояснення з використанням станів та структури інтелектуальної системи

Підхід	Особливості
1. На основі спрощення процесу функціонування інтелектуальної системи	Для пояснення використовуються моделі «поверхневих» знань.
2. На основі виділення впливу окремих функцій та дій на кінцевий результат	В рамках даного підходу в якості пояснення відображається вплив окремих етапів процесу прийняття рішення або окремих функцій, що були використані у даному процесі, на показники отриманого результату.
3. На основі відомих концепцій у предметній області	Згідно даного підходу виконується візуалізація внутрішнього стану моделі, що лежить в основі процесу прийняття рішення.

Тобто складні закономірності процесу прийняття рішення, отримані, наприклад, в результаті машинного навчання, представляються спрощеними причинно-наслідковими або темпоральними відношеннями, які задають зв’язок між ключовими подіями та фактами предметної області, без деталізації такого зв’язку.

Другий підхід до формування пояснень відображає вплив окремих дій або функцій на кількісні властивості отриманого в інтелектуальній системі результат. В якості показників використовуються, наприклад, точність, AUC, тощо.

При реалізації даного підходу застосовуються, наприклад, діаграма середнього зниження точності (Mean Decrease Accuracy – MDA), метод DeepLIFT.

MDA показує, наскільки зменшується точність за умови виключення відповідної змінної із моделі процесу прийняття рішення в інтелектуальній системі. Змінні на діаграмі представляються за зменшенням їх важливості.

Метод DeepLIFT виконує декомпозицію результату роботи нейронної мережі для конкретного входу за рахунок зворотного розповсюдження відкликів нейронів на кожну ознаку вхідного сигналу. Даний метод надає оцінку вкладу кожного нейрону у результат, порівнюючи поточну та еталонну активації.

Підхід до побудови пояснень на основі концепцій та понять орієнтований на те, щоб показати вплив окремих структурних елементів моделі процесу прийняття рішення в інтелектуальній системі на кінцевий результат. Для відображення даного впливу використовується кількісна оцінка впливу певних понять на отримане в системі рішення. Тобто вхідні дані сортуються за їх відповідністю певним поняттям та концепціям, після чого встановлюється вплив цих понять. Наприклад, вхідні зображення для класифікації можуть бути відсортовані за кольором, за формою ліній на зображенні (точкова, зигзагами, тощо), за об’єктами на зображенні (чоловік, жінка, дитина).

При використанні нейронних мереж даний підхід реалізується методом кількісного тестування з векторами активації концепції (Testing with Concept Activation Vectors – TCAV). Метод створює орієнтовану на розуміння людиною лінійну інтерпретацію внутрішніх станів моделі глибокого навчання в термінах понять та концепцій предметної області.

За другим напрямком виділяються такі підходи:

– візуалізація впливу вхідних даних на отримане рішення;

– використання прикладів прийняття рішення.

Ключові відмінності даних підходів наведено у табл. 2.

Таблиця 2 – Пояснення з використанням вхідних даних

Підхід	Особливості
1. Візуалізація впливу вхідних даних	Даний підхід орієнтований на представлення графіків, що відображають деталізовані або усереднені залежності результату від вхідної функції.
2. На основі типових прикладів роботи системи	Задаються приклади, що найкращим чином ілюструють процес прийняття рішення в інтелектуальній системі.

При візуалізації пояснення використовуються, як правило, графіки, що відображають індивідуальне умовне очікування (Individual Conditional Expectation – ICE), або ж часткові залежності (Partial Dependence Plots – PDPs).

В першому випадку на графіку представляється задається індивідуальна залежність певного показника від заданого параметра або функції (наприклад, залежність захворюваності від віку людини). В другому випадку задається середнє значення цього показника.

Підхід до побудови пояснень на основі прикладів передбачає побудову множини прикладів, що ілюструють процес отримання результатів в інтелектуальній системі.

Локальний, груповий та глобальний рівні деталізації пояснень. Розглянуті у попередньому підрозділі підходи орієнтовані на побудову пояснення на локальному, груповому та глобальному рівнях деталізації.

Локальне пояснення призначено для обґрунтування користувачеві одного конкретного результату.

Групове пояснення забезпечує розуміння користувачем підмножини схожих результатів роботи інтелектуальної системи.

Глобальне пояснення робить «прозорим» безпосередньо процес отримання рішення в інтелектуальній системі.

Між цими рівнями деталізації можна встановити такий зв'язок. Пояснення глобального рівня відображає послідовність впливів окремих дій або функцій процесу формування рішення в інтелектуальній системі на кінцевий результат.

Процес формування рішення реалізується декілька разів для отримання різних результатів. Тому глобальне пояснення містить залежності, що є спільними для всіх цих екземплярів процесу отримання результату.

Групове пояснення містить залежності для підмножини екземплярів процесу отримання рішення. Тому в даному випадку використовується підмножина залежностей, типи яких відрізняються від узагальнених залежностей глобального пояснення. Наприклад, обмеження для групового пояснення не завжди є обмеженнями для повного процесу прийняття рішення.

Локальні пояснення у даній ієрархії доцільно розглядати як залежності між конкретним набором вхідних даних та отриманим в інтелектуальній системі результатом. Відмінності між вказаними рівнями пояснень наведено на рис.2.

Локальне пояснення надається для одного екземпляру $E_i = \langle e_{i,1}, e_{i,2}, ..., e_{i,J} \rangle$ процесу прийняття рішення та пов'язує першу $e_{i,1}$ та останню $e_{i,J}$ події, що відображають виконання цього процесу.

Тобто локальне пояснення обґрунтовує кінцевий результат для користувача, задаючи зв'язок між вхідними даними та кінцевим результатом $\pi_{i,J}^{i,1}$ у вигляді функціональної залежності $f(e_{i,1}, e_{i,J})$ між подією отримання вхідних даних $e_{i,1}$ та результату $e_{i,J}$:

$$\pi_{i,J}^{i,1} \equiv f(e_{i,1}, e_{i,J}). \quad (1)$$

Групове пояснення містить залежності для підмножини $E^{\text{group}} = \{E_i\}$, $E^{\text{group}} \subset E$. множини E всіх можливих реалізацій процесу прийняття рішення в інтелектуальній системі. Групове пояснення має вигляд:

$$\begin{aligned} &\pi g_m^j \equiv f(e_j, e_m) \,\big|\, \exists E^{\text{group}} = \{E_i\} : \\ &(\exists i) : e_{i,j} \equiv e_j, e_{i,m} \equiv e_m. \end{aligned} \quad (2)$$

Відповідно до (2), групове пояснення містить у собі в якості окремого елемента локальне пояснення в тому випадку, якщо E^{group} містить лише один елемент E_1:

$$\pi_{i,J}^{i,1} \equiv \pi g_m^j \,\big|\, E^{\text{group}} = \{E_1\},\ j = 1, J = m. \quad (3)$$

Глобальне пояснення π_m^j об'єднує всі відомі реалізації процесу отримання рішення в інтелектуальній інформаційній системі:

$$\begin{aligned} &\pi_m^j \equiv f(e_j, e_m) : \\ &1. \exists i : e_{i,j} \equiv e_j, e_{i,m} \equiv e_m, \\ &2. \exists j \exists m : (\forall i) e_{i,j} \equiv e_j, e_{i,m} \equiv e_m. \end{aligned} \quad (4)$$

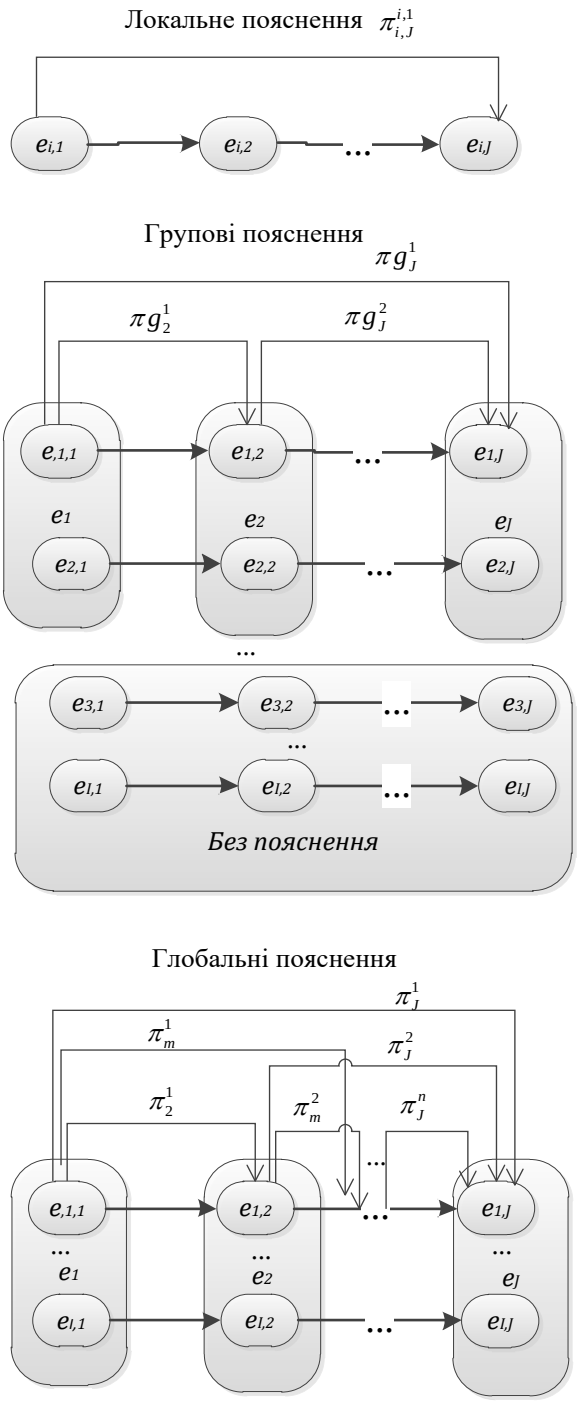

Рис. 2. Локальні, групові та глобальні пояснення

Глобальне пояснення містить залежності у формі умов та обмежень виконання дій процесу прийняття рішення.

Умови (1) визначають, що представлена поясненням залежність є дійсною лише для підмножини реалізацій процесу формування результату. Тому умова має бути актуальною для конкретної реалізації, для

якої застосовується глобальне пояснення. Позначимо умову як πr_m^j.

Тоді πr_m^j може бути використана для пояснення таких E_i реалізацій процесу отримання результату у інтелектуальній системі, для яких виконується:

$$\left(\forall E_i\right)\pi r_m^j \equiv \pi_{i,m}^{i,j}. \quad (5)$$

Обмеження виконується для всіх реалізацій процесу прийняття рішення і тому обмеження як глобальне пояснення може бути використано у всіх випадках отримання результату.

Умова (5) фактично відповідає локальному поясненню, тобто:

$$\left(\forall E_i \in E^{\text{group}}\right)\pi r_m^j \equiv \pi g_m^j. \quad (6)$$

Таким чином, глобальні пояснення з формальної точки зору поєднують пояснення на основі обмежень виконання дій з отримання результату або відповідних функцій, а також групові та локальні пояснення.

Тому в подальшому будуть розглядатись лише глобальні пояснення π_m^j.

Послідовності глобальних пояснень, наприклад $\Pi' = \left\langle \pi_2^1, \pi_3^2, \pi_I^{I-1} \right\rangle$ або $\Pi'' = \left\langle \pi_2^1, \pi_3^1, ..., \pi_I^1 \right\rangle$ відображають процес прийняття рішення у формі, прозорій для користувача.

Однак з цієї причини пояснення не в повній мірі відображає стани інтелектуальної системи і процес прийняття рішень в цілому. Тому модель пояснення має враховувати стохастичну складову при описі системи, щодо якої формується пояснення, а також можливі зміни ефективності роботи системи в результаті коригування взаємодії з користувачем після отримання пояснення.

Відповідно, при формуванні послідовності пояснень необхідно оцінювати їх ефективність. Така оцінка ефективності може бути в подальшому представлена вагою пояснення w_m^j, що враховує його стохастичний аспект. Виділення ваги дає можливість відібрати кращі за цим показником пояснення із множини можливих.

Оцінка може бути виконана для кожного пояснення в послідовності або ж для всієї послідовності пояснень. В другому випадку може бути виконано формування та порівняння декількох послідовностей пояснень з подальшою динамічною зміною цих послідовностей в залежності від результатів взаємодії з користувачем.

Модель пояснення на локальному, груповому та глобальному рівнях деталізації. Пояснення згідно (4) встановлюють зв'язок між парами подій, що відображають стани інтелектуальної системи. Тобто наступний стан у поясненні залежить від попереднього стану і пояснювальної дії, яка привела до переходів між цими станами. Тобто ми маємо відповідність між станами пояснення s_j та подіями e_j, що відображають процес прийняття рішення. Відповідно, пояснення може бути представлено у вигляді:

$$\pi_m^j \equiv f\left(s_j, s_m\right) \left| \exists e_j \mapsto s_j, \right| \exists e_m \mapsto s_m. \quad (7)$$

Тоді для вибору найкращого пояснення необхідно оцінити його ефективність. Як було показано вище, локальні та групові пояснення є окремим випадком глобального пояснення. Глобальне ж пояснення містить множину можливих елементарних пояснень π_m^j, що складаються у послідовність Π, яка забезпечує прозорість всього процесу від початкової події e_1 і до кінцевої події e_J. Це свідчить про важливість використати оцінку всього процесу. У випадку кінцевої множини подій процесу, для якого надається рішення, така оцінка матиме вигляд суми (8) або середнього значення (9) оцінок w_m^j окремих пояснень π_m^j.

$$W_\Pi^{'} = \sum_{j,m} w_m^j. \quad (8)$$

$$W_\Pi^{''} = \frac{\sum_{j,m} w_m^j}{\left|\left\{w_{j,m}\right\}\right|}. \quad (9)$$

Слід зазначити, що внаслідок стохастичних характеристик процесу пояснень оцінка окремого пояснення може бути обчислена на основі ймовірності використання послідовності пояснень в цілому.

У випадку, якщо кількість подій є невідомою (наприклад, для ітеративного процесу формування рішення у взаємодії з користувачем), така оцінка має враховувати в першу чергу важливість поточних пояснень. Для цього зазначена оцінка використовує коефіцієнт γ, що лежить в діапазоні від нуля до 1:

$$W_\Pi^{'''} = \sum_{j,m} \frac{\gamma^{j-1} + \gamma^{m-1}}{2} w_m^j. \quad (10)$$

Таким чином, модель послідовності глобальних пояснень має вигляд:

$$M = \left\{\Pi, W_\Pi\right\}, \quad (11)$$

де $\Pi = \left\langle \pi_m^1, ..., \pi_m^j, ..., \pi_I^{I-1} \right\rangle$, а W_Π обчислюється згідно (8)–(10).

Дана модель об'єднує, як було показано у (3) та (6), локальні та групові пояснення.

Висновки. Запропоновано узагальнену модель пояснення для локального, групового та глобального рівнів деталізації, яка представлено упорядкованою послідовністю зважених залежностей між подіями або станами процесу прийняття рішення в інтелектуальній системі, що дає можливість в рамках глобального пояснення представити локальне пояснення на основі ланцюжку групових пояснень між подіями отримання вхідних даних та результуючого рішення.

У практичному аспекті розроблена модель орієнтована на використання розглянутих підходів побудови пояснень на основі спрощення процесу функціонування інтелектуальної системи та на основі виділення впливу окремих функцій та дій на кінцевий

результат. Деталізація подій процесу прийняття рішення з урахуванням змінних, які характеризують ці події, дає можливість використати підхід до побудови пояснень на основі відомих концепцій та понять у предметній області.

Список літератури

1. Engelbrecht Andries P. *Computational Intelligence: An Introduction*. New York: John Wiley & Sons, 2007. 632 p.
2. Castelvecchi D. Can we open the black box of AI? *Nature News*. 2016. Vol. 538 (7623). P. 20–23.
3. Gunning D., Aha D. DARPA's Explainable Artificial Intelligence (XAI) Program. *AI Magazine*. 2019. Vol. 40 (2). P. 44–58.
4. Preece A., Harborne D., Braines D., Tomsett R., Chakraborty S. Stakeholders in Explainable AI. *arXiv:1810.00184*. 2018.
5. Gilpin L. H., Bau D., Yuan B. Z., Bajwa A., Specter M., Kagal L. Explaining Explanations: An Overview of Interpretability of Machine Learning. *arXiv:1806.00069*. 2018.
6. Miller T. Explanation in artificial intelligence: Insights from the social sciences, *Artificial. Intelligence*. 2019. Vol. 267 P. 1–38.
7. Zhang Q., Wu N. Y., Zhu S.-C. Interpretable convolutional neural networks, *IEEE Conference on Computer Vision and Pattern Recognition*. 2018. P. 8827–8836.
8. Deng H. Interpreting tree ensembles with intrees. *arXiv:1408.5456*. 2014.
9. Чалий С.Ф., Лещинський В.О., Лещинська І.О. Декларативно-темпоральний підхід до побудови пояснень в інтелектуальних інформаційних системах. *Вісник Нац. техн. ун-ту "ХПІ": зб. наук. пр. Темат. вип. Системний аналіз, управління та інформаційні технології*. Харків: НТУ «ХПІ». 2020. № 2 (4). С. 51–56.
10. Halpern J. Y., Pearl J. Causes and explanations: A structural-model approach. Part I: Causes. *The British Journal for the Philosophy of Science*. 2005. № 56 (4). P. 843–887.
11. Chalyi S., Leshchynskyi V. Temporal representation of causality in the construction of explanations in intelligent systems. *Advanced Information Systems*. Kharkiv: NTU "KhPI", 2020. Vol. 4, № 3. P. 113–117.
12. Чалий С.Ф., Лещинський В.О., Лещинська І.О. Моделювання пояснень щодо рекомендованого переліку об'єктів з урахуванням темпорального аспекту вибору користувача. *Системи управління, навігації та зв'язку*. 2019. Том 6, № 58. С. 97–101.

References (transliterated)

1. Engelbrecht Andries P. *Computational Intelligence: An Introduction*. NJ, John Wiley & Sons, 2007. 632 p.
2. Castelvecchi D. Can we open the black box of AI? *Nature News*. 2016, vol. 538 (7623), pp. 20–23/
3. Gunning D., Aha D. DARPA's Explainable Artificial Intelligence (XAI) Program. *AI Magazine*. 2019, no 40 (2), pp. 44–58.
4. Preece A., Harborne D., Braines D., Tomsett R., Chakraborty S. Stakeholders in Explainable AI. *arXiv:1810.00184*. 2018.
5. Gilpin L. H., Bau D., Yuan B. Z., Bajwa A., Specter M., Kagal L. Explaining Explanations: An Overview of Interpretability of Machine Learning. *arXiv:1806.00069*. 2018.
6. Miller T. Explanation in artificial intelligence: Insights from the social sciences, *Artificial*. *Intelligence*. 2019, vol. 267, pp. 1–38.
7. Zhang Q., Wu N. Y., Zhu S.-C, Interpretable convolutional neural networks, *IEEE Conference on Computer Vision and Pattern Recognition*. 2018, pp. 8827–8836.
8. Deng H. Interpreting tree ensembles with intrees. *arXiv:1408.5456*. 2014.
9. Chalyi S., Leshchynskyi V., Leshchynska I. Deklaratyvno-temporalnyi pidkhid do pobudovy poiasnen v intelektualnykh informatsiinykh systemakh [Declarative-temporal approach to the construction of explanations in intelligent information systems]. *Visnyk Nats. tekhn. un-tu "KhPI": zb. nauk. pr. Temat. vyp. Systemnyi analiz, upravlinnia ta informatsiini tekhnolohii* [Bulletin of the National Technical University "KhPI": a collection of scientific papers. Thematic issue: System analysis, management and information technology]. Kharkov, NTU "KhPI" Publ, 2020, no. 2(4), pp. 51–56.
10. Halpern J. Y., Pearl J. Causes and explanations: A structural-model approach. Part I: Causes. *The British Journal for the Philosophy of Science*. 2005, no. 56 (4), pp. 843–887.
11. Chalyi S., Leshchynskyi V. Temporal representation of causality in the construction of explanations in intelligent systems. *Advanced Information Systems*. 2020, vol. 4, no 3, pp. 113–117.
12. Chalyi S. F., Leshchynskyi V. O., Leshchynska I. O. Modelyuvannya poyasnen shodo rekomendovanogo pereliku ob'yektiv z urahuvannyam temporalnogo aspektu viboru koristuvacha [Modeling explanations for the recommended list of items based on the temporal dimension of user choice]. *Sistemi upravlinnya, navigaciyi ta zv'yazku* [Control, Navigation and Communication Systems]. 2019, vol. 6, no 58, pp. 97–101.

Надійшла (received) 01.11.2022

Відомості про авторів / About the Authors

Чалий Сергій Федорович – доктор технічних наук, професор, Харківський національний університет радіоелектроніки, професор кафедри інформаційних управляючих систем, м. Харків, Україна; ORCID: https://orcid.org/0000-0002-9982-9091; e-mail: serhii.chalyi@nure.ua

Лещинський Володимир Олександрович – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри програмної інженерії, м. Харків, Україна; ORCID: https://orcid.org/0000-0002-8690-5702; e-mail: volodymyr.leshchynskyi@nure.ua

Chalyi Serhii Fedorovych – Doctor of Technical Sciences, Professor, Kharkiv National University of Radio Electronics, Professor of the Department of Information Control System, Kharkiv; ORCID: https://orcid.org/0000-0002-9982-9091; e-mail: serhii.chalyi@nure.ua

Leshchynskyi Volodymyr Oleksandrovich – PhD, Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor of the Department of Software Engineering, Kharkiv; ORCID: https://orcid.org/0000-0002-8690-5702; e-mail: volodymyr.leshchynskyi@nure.ua