

*С.Ф. ЧАЛИЙ, В.О. ЛЕЩИНСЬКИЙ*

## **ТЕМПОРАЛЬНО-КАУЗАЛЬНІ МЕТОДИ ПОБУДОВИ ПОЯСНЕНЬ В СИСТЕМАХ ШТУЧНОГО ІНТЕЛЕКТУ**

Предметом дослідження є процес побудови пояснень в системах штучного інтелекту. Запропоновано темпорально-каузальний підхід до побудови пояснень, який структурує представлення пояснень на локальному, проміжному та глобальному рівнях у темпоральному та каузальному аспектах. Розроблено узагальнений метод побудови пояснень з використанням темпоральних та каузальних залежностей, а також метод уточнення пояснень з використанням темпоральних залежностей. Крім того, запропоновано метод побудови пояснень на глобальному рівні на основі упорядкованості вхідних даних, що враховує структуру даних при побудові пояснення.

### **1. Вступ**

Системи штучного інтелекту (ШІ) широко застосовуються в таких сферах людської діяльності, як охорона здоров'я, фінанси, менеджмент, освіта тощо. Вони забезпечують підтримку прийняття рішень, які мають суттєвий вплив на життя людей. Однак притаманна моделям ШІ непрозорість створює суттєві перешкоди для довіри користувачів до їх рішень та, відповідно, зменшує ефективність використання таких систем. З метою вирішення даної проблеми проводяться дослідження в галузі пояснюваного ШІ (Explainable AI) [1]. Пояснення мають забезпечити прозорість систем ШІ за рахунок безпосередньої інтерпретації процесу прийняття рішень в таких системах [2]. Побудова релевантних пояснень в системах ШІ потребує визначення залежностей, що привели до рішення системи [3]. Ці залежності базуються на відображенні процесу прийняття рішень в часі, тобто на послідовності дій та подій, що привели до конкретного результату у системі ШІ [4]. Упорядкованість цих дій у часі представляється темпоральними залежностями [5]. Каузальні залежності розкривають причинно-наслідкові зв'язки між вхідними даними, проміжними станами системи та рішенням системи, створюючи умови для розуміння механізмів, що лежать в основі функціонування моделі ШІ [6]. Тому поєднання темпорального та каузального опису процесу формування рішення дає можливість суттєво підвищити його інтерпретованість та, як наслідок, довіру до рішення з боку користувачів. Такі пояснення дають можливість простежити ланцюжок подій щодо формування рішення з урахуванням причинно-наслідкових зв'язків та темпоральних зав'язків між ними. Це, в свою чергу, надає користувачам можливість оцінити рішення системи, виявити потенційні упередження та адаптувати рішення відповідно до своїх потреб. Таким чином, проблема побудови пояснень в системах ШІ з урахуванням темпорального та каузального аспектів є актуальною [7]. Вирішення цієї проблеми створює умови для побудови прозорого процесного опису механізму прийняття рішень в інтелектуальних системах.

### **2. Аналіз літературних даних і постановка проблеми дослідження**

Дослідження, що проводяться в галузі пояснюваного ШІ (XAI) [8], направлені на розробку систем, алгоритми роботи яких можуть бути безпосередньо інтерпретовані, що забезпечує прозорість процесу прийняття рішень [9], [10]. Ключові підходи, що сьогодні застосовуються на практиці при побудові пояснень, направлені на виявлення основних факторів, які вплинули на отримане рішення. Такі фактори визначаються через побудову спрощеної локальної моделі процесу прийняття рішення [11] або через визначення внеску кожної ознаки у передбачення системи на основі підходу, прийнятого в теорії ігор [12]. Однак при побудові пояснень не завжди враховуються важливість відображення процесу

прийняття рішень одночасно у двох аспектах: у темпоральному – для визначення типових послідовностей дій із прийняття рішення та у каузальному – для визначення причин цих дій. Врахування обох аспектів створює умови для індивідуалізації пояснення та дає можливість зробити зрозумілим процес прийняття рішень в системі ШІ. Темпоральні та каузальні аспекти пояснень розглядаються в низці досліджень. Зокрема, в [6], [13] представлено підходи до моделювання альтернативних причин рішення на основі темпоральної впорядкованості подій та формування каузальних залежностей з урахуванням такої впорядкованості. Темпоральна впорядкованість процесу прийняття рішення може бути представлена у вигляді множини темпоральних правил [14]–[16], з яких можна сформулювати можливі альтернативи цього процесу. Використання каузальних залежностей для побудови пояснень розглянуто в роботах [17], [18]. В [19] показано, що люди виявляють закономірності у процесах та явищах, розглядаючи їх альтернативні варіанти.

Результати зазначених досліджень обумовлюють можливість побудови каузальних пояснень для розуміння людьми процесу прийняття рішення та результуючого рішення системи ШІ. Однак побудові підходу, який би формував пояснення з урахуванням одночасно каузального та темпорального аспектів, не приділялось достатньо уваги, що свідчить про актуальність теми даного дослідження.

### **3. Мета і задачі дослідження**

Метою дослідження є розробка темпорально-каузального підходу до побудови пояснень в системах ШІ з тим, щоб представити пояснення як щодо процесу прийняття рішення, так і щодо отриманого рішення та зробити їх прозорими та зрозумілими для вирішення практичних завдань користувачів.

Для досягнення поставленої мети необхідно вирішити такі задачі: структуризація рівнів представлення пояснень з урахуванням темпорального й каузального аспектів; розробка узагальненого методу побудови пояснень з використанням темпоральних та каузальних залежностей; розробка методу уточнення пояснень з використанням темпоральних залежностей.

### **4. Структуризація рівнів представлення пояснень з урахуванням темпорального й каузального аспектів**

Представлення пояснення як щодо процесу, так і щодо результуючого рішення системи ШІ пов'язано із потребами внутрішніх та зовнішніх користувачів інтелектуальної системи. Внутрішні та зовнішні користувачі мають різні вимоги щодо прозорості процесу прийняття рішень в системі ШІ. Ці відмінності обумовлені специфікою завдань, які вирішують користувачі, а також їхнім рівнем технічної експертизи та знаннями предметної області. Зовнішні користувачі, як правило, використовують рішення, сформовані системами ШІ. Вони використовують ці результати для підтримки своєї діяльності. Зовнішні користувачі висувають такі вимоги до рішень систем ШІ, які мають суттєвий вплив на характеристики пояснень:

– зрозумілість та інтерпретованість результатів: рішення інтелектуальної системи мають бути представлені у формі, доступній для розуміння користувачем, з урахуванням його рівня знань та контексту використання;

– довіра до системи: користувачі мають бути впевнені у надійності та обґрунтованості рішень інтелектуальної системи, що передбачає надання пояснень щодо причин та факторів, які вплинули на формування поточного результату;

— можливість уточнення результату: зовнішні користувачі повинні мати можливість коригувати вхідні дані або параметри системи з тим, щоб отримати рішення, що відповідають їх потребам та перевагам.

Вимоги внутрішніх користувачів систем ШІ, які розробляють та супроводжують такі

системи, містять в собі:

– прозорість та інтерпретованість моделей прийняття рішення: внутрішні користувачі потребують пояснень щодо внутрішньої логіки прийняття рішень системи ШІ з тим, щоб перевірити її відповідність тим задачам, що вирішує система;

– можливість налагодження та удосконалення: пояснення мають створювати умови для виявлення та усунення помилок і вузьких місць процесу прийняття рішення, а також для підвищення ефективності моделі шляхом налаштування її параметрів;

забезпечення відповідності регуляторним вимогам: внутрішні користувачі мають впевнитись, що система ШІ формує рішення у відповідності до етичних норм, а також вимог до безпеки та конфіденційності даних.

Таким чином, пояснення для користувачів мають бути двох типів:

– процесні пояснення, що розкривають послідовність кроків із формування рішення, а також залежності між цими кроками;

– реконструктивні пояснення, що відтворюють ключові причини рішення системи ШІ, тобто надають користувачеві інформацію про те, як конкретні вхідні дані або фактори вплинули на фінальний результат.

Процесні пояснення орієнтовані в першу чергу на внутрішніх користувачів і тому мають надавати детальнішу інформацію щодо процесу прийняття рішень. Це дає можливість виявляти «вузькі місця» та можливі напрямки вдосконалення роботи системи.

Реконструктивні пояснення призначені переважно для зовнішніх користувачів. Тому вони можуть використовувати спрощені моделі, що визначають залежність результату від найважливіших вхідних змінних. Такі пояснення мають сприяти підвищенню довіри користувачів до системи ШІ.

Ключові властивості процесних та реконструктивних пояснень представлено у таблиці 1.

Таблиця 1

Представлення пояснень у темпоральному та каузальному аспектах

Тип пояснень	Властивості
Процесні	– залежності між діями процесу прийняття рішення; – залежності між діями процесу прийняття рішення та отриманим результатом; – пояснення представляються у темпоральному та каузальному аспекті.
Реконструктивні	– каузальні залежності встановлюються між вхідними даними та отриманим результатом; – темпоральний аспект може бути врахований як упорядкованість вхідних даних.

Узагальнені результати структуризації пояснень у темпоральному та каузальному аспектах наведено в таблиці 2.

Структуризація пояснень у темпоральному та каузальному аспектах з урахуванням відмінностей процесних та реконструктивних типів пояснень дає можливість врахувати різницю у вимогах внутрішніх та зовнішніх користувачів.

Система ШІ може бути представлена у вигляді «білої», «сірої» або «чорної» скриньки. У випадку представлення системи як «біла скринька» підсистема пояснень розробляється разом із системою ШІ або ж використовуються безпосередньо інтерпретовані алгоритми (правила, дерева рішень). Задача побудови пояснень в даному випадку зводиться до представлення користувачеві відомих залежностей, що були

Таблиця 2

Представлення пояснень у темпоральному та каузальному аспектах

Рівень представлення	Властивості
Локальний	– темпоральні залежності між парами дій або станів процесу прийняття рішення; – каузальні обмеження щодо послідовності пар дій у процесі прийняття рішення.
Проміжний	– темпоральні залежності, що відображають вплив дій процесу прийняття рішення на результат роботи системи штучного інтелекту; – каузальні залежності, що визначають зв'язок між діями процесу а також рішенням системи.
Глобальний	– каузальні залежності, що відображають ключові вхідні дані, значення яких представлені як причина рішення системи штучного інтелекту; – каузальні залежності, що відображають вплив структури вхідних даних на рішення системи.

використані у процесі отримання рішення. Коли система ШІ представлена як «сіра» або «чорна скринька», підсистема пояснень має доповнювати існуючу систему та формувати пояснення в умовах невизначеності внаслідок обмеженого або відсутнього доступу до інформації щодо проміжних станів цієї системи. Така ситуація може виникати внаслідок використання юридичних механізмів захисту інтелектуальної власності (система як «чорна скринька») або застосування непрозорих моделей машинного навчання (система представляється переважно як «сіра скринька»). В першому випадку пояснення відображають статичні характеристики процесу прийняття рішень, оскільки інформація про послідовність дій не розглядається. У другому випадку інформація про процес прийняття рішення є частково доступною, що дозволяє будувати пояснення, які відображають темпоральний та каузальний аспект даного процесу.

Таким чином, задача побудови пояснень підрозділяється на такі підзадачі:

- побудова пояснень для внутрішніх користувачів у формі правил, що відображають упорядкованість дій процесу прийняття рішення в системі ШІ;
- побудова пояснень для внутрішніх користувачів у формі каузальних залежностей, що визначають причини дій процесу прийняття рішення;
- побудова пояснень для зовнішніх користувачів у вигляді каузальних залежностей, що відображають вхідні дані як причини отриманого рішення.

## 5. Темпорально-каузальний метод побудови пояснень

### 5.1. Основні етапи методу

Метод побудови пояснень, який формує пояснення на розглянутих рівнях представлення, містить такі етапи.

Етап 1. Побудова темпоральних залежностей та каузальних обмежень щодо пар дій процесу прийняття рішення на локальному рівні деталізації пояснень.

Крок 1.1. Побудова темпоральних залежностей для упорядкованих у часі пар дій або станів  $s_{i,j}$  та  $s_{i,j+n}$   $j$ -го процесу прийняття рішення. Результатом кроку є множина  $\{\{s_{i,j}, s_{i,j+n}\}\}$ , для якої виконується умова  $s_{i,j} \rightarrow s_{i,j+n}$ , тобто стан  $s_{i,j}$  був досягнутий у часі раніше, ніж стан  $s_{i,j+n}$ .

Крок 1.2. Формування обмежень  $R(s_{i,j}, s_{i,j+n}) \equiv R_{j+n}^j$  на послідовність дій.

Пара станів є обмеженням за умови:

$$R_{j+n}^j = \text{true iff } (\forall i) s_{i,j} \rightarrow s_{i,j+n}. \quad (1)$$

Згідно з (1), обмеження  $R_{j+n}^j$  має виконуватись для всіх відомих реалізацій процесу прийняття рішення. Тоді обмеження  $R_{j+n}^j$  в рамках пояснення визначає стан  $s_{i,j}$  причиною стану  $s_{i,j+n}$  (або відповідної дії, що привела до даного стану).

Відповідно, дане обмеження в рамках пояснення може розглядатись як каузальна залежність  $C_{j+n}^j$ :

$$R_{j+n}^j \equiv C_{j+n}^j. \quad (2)$$

Тобто, згідно з таблицею 2, на даному етапі формуються темпоральні та каузальні залежності.

Етап 2. Побудова темпоральних правил та каузальних залежностей щодо впливу дій процесу прийняття рішення на отриманий результат.

Крок 2.1. Побудова темпоральних залежностей  $\langle s_{i,j}, y_i \rangle$  між станом  $s_{i,j}$  (або відповідною дією) та поточним рішенням системи  $y_i$ .

Крок 2.2. Формування темпоральних правил  $F_j$  шляхом узагальнення темпоральних залежностей  $\langle s_{i,j}, y_i \rangle$  та темпоральних правил  $X_{j,j+n}$  шляхом узагальнення темпоральних залежностей  $\langle s_{i,j}, s_{i,j+n} \rangle$ .

Крок 2.3. Розрахунок достатніх умов щодо правил  $F_j$  та  $X_{j,j+n}$  для побудови каузальних залежностей.

Достатня умова може бути розрахована з використанням показників ймовірності або можливості застосування правила. Показник ймовірності доцільно використовувати в тому випадку, якщо багаторазове формування рішення виконувалось в однакових умовах, тобто не використовувалось додаткове налагодження відповідного процесу. В іншому випадку для розрахунків використовується показник можливості, оскільки він дає можливість оперувати довільними числовими оцінками змінних. Показники можливості розраховуються згідно з теорією можливості. Для обчислення можливості каузальної залежності на даному етапі може також бути використані дані щодо ваг антецедента та консеквента правила. Такий підхід доцільно використовувати у випадку, якщо ці ваги враховують специфіку предметної області. Наприклад, відображають значимість складових правила для отриманого в системі ШІ результату.

Крок 2.4. Розрахунок необхідних умов щодо правил  $F_j$  та  $X_{j,j+n}$  для побудови каузальних залежностей.

Необхідні умови визначаються у випадку використання можливісної оцінки на попередньому кроці. Тоді на даному кроці розраховується показник необхідності. Обчислення даного показника виконується згідно з положеннями теорії можливості.

Крок 2.5. Формування каузальних залежностей.

На даному кроці темпоральні правила з максимальним значенням можливості за

умови перевищення порогового значення необхідності розглядаються як каузальні залежності.

Етап 3. Побудова каузальних залежностей щодо впливу вхідних змінних на рішення системи ШІ.

Крок 3.1. Розрахунок показника можливості для всіх значень всіх вхідних змінних.

Крок 3.2. Розрахунок показника необхідності для всіх значень всіх вхідних змінних.

Крок 3.3. Відбір змінних за найбільшим значенням можливості при перевищенні порогового значення необхідності.

Крок 3.4. Формування каузальних залежностей для відібраних вхідних змінних.

Результатом методу є темпоральні та каузальні залежності для внутрішніх користувачів, які пояснюють вплив окремих дій на інші дії або послідовностей дій на рішення системи. Результатом також є каузальні залежності, які пояснюють вплив значень і упорядкованості вхідних даних на рішення системи ШІ для зовнішніх користувачів.

## **5.2. Приклад застосування методу**

Розглянемо приклад застосування даного методу для пояснення зовнішньому користувачеві щодо рекомендованого в системі електронної комерції ноутбуку з використанням можливісних оцінок. В даному випадку використовується інформація про ключові компоненти ноутбуку (процесор, пам'ять тощо), щоб пояснити, чому було рекомендовано саме цей комп'ютер. На кроці 3.1 розраховується показник можливості для всіх моделей процесора, пам'яті тощо, які використовувались у ноутбуках даної цінової категорії. Показник можливості розраховується, зокрема, на основі відсотка ноутбуків із поточною моделлю процесора. Показник ймовірності в даному випадку використовувати недоцільно, оскільки протягом періоду, що аналізується, партії комп'ютерів з різними моделями процесів могли надходити в довільному порядку. Кількість комплектуючих певного типу також могла бути обмежена.

Після розрахунку показника необхідності за вхідними змінними (модель процесора, об'єм пам'яті тощо) відбираються значення змінних з максимальною оцінкою можливості за умови порогового значення необхідності (крок 3.3). Ці значення сортуються за показником можливості. Змінна з найбільшим значенням показника подається в поясненні як ключова причина рекомендації ноутбука в системі електронної комерції. Тобто якщо конкретна модель процесора *i7-1165* має найбільше значення показника можливості, то ця модель розглядатиметься як основна причина рекомендації ноутбука з таким процесором.

## **6. Метод побудови пояснень на основі упорядкованості вхідних даних**

### **6.1. Основні етапи методу**

Даний метод побудови пояснень являє собою імплементацію етапу 3 описаного вище темпорально-каузального методу побудови пояснень з використанням заданих ваг антецедента та консеквента правил. Його відмінність полягає в тому, що він використовує темпоральні залежності для опису вхідних даних, тобто комбінує темпоральний і каузальні аспекти на рівні вхідних даних. Вхідними даними методу є ваги станів  $s_{i,j}$  за умови, що ці ваги відображають особливості предметної області, що є суттєвими для побудови пояснень.

Метод містить такі ключові етапи.

Етап 1. Побудова темпоральних залежностей, що відображають упорядкованість вхідних даних. В даному випадку враховується той факт, що темпоральні правила не використовують абсолютне значення часу, а відображають відносну упорядкованість подій або станів у часі. Тому вони можуть бути використані для відображення відносної упорядкованості вхідних даних.

Етап 2. Розрахунок ваг отриманих залежностей на основі ваг станів та формування зважених каузальних залежностей.

На даному етапі підсумовуються ваги станів.

Етап 3. Формування можливих підмножин зважених каузальних залежностей, що описують вхідну послідовність даних.

Етап 4. Розрахунок ваг отриманих підмножин як нормалізованої суми ваг правил.

Етап 5. Упорядкування підмножин за вагами.

Етап 6. Формування набору пояснень із підмножин правил.

## 6.2. Експериментальна перевірка методу

Експериментальну перевірку виконано для пояснення результатів класифікації спаму з використанням нейромережі. Остання, з точки зору кінцевого користувача, являє собою «чорну скриньку». Як вхідні дані були використані спамові повідомлення із зваженими словами. Ваги слів у вхідному повідомленні відображають їх відповідність спаму. Тобто позитивна вага слова свідчить, що це слово може бути складовою спаму. Проте порівняльний аналіз ваг слів потребує значних витрат часу кваліфікованого спеціаліста на формування однозначного пояснення.

Результати побудови набору пояснень на основі каузальних залежностей для вхідного повідомлення «Free camera phones with line rental from 4/month. 1.2 price txt bundle deals available» представлено на рисунку 1.

```
i1: [(0,1), (0,16), (0,18), (1,16), (1,18), (16,18)] | w=2.1895959619532865
i2: [(0,1), (0,16), (1,16), (1,18), (16,18)] | w=2.032710511557881
i3: [(0,1), (0,16), (1,16), (16,18)] | w=1.858851006878812
i4: [(0,16), (1,16), (16,18)] | w=1.5830576336513666
i5: [(0,16), (1,16)] | w=1.1289856862277166
i6: [(1,16)] | w=0.5729798702556901
```

index	text
0	FREE
1	camera
16	txt
18	deals

Рис. 1. Набори правил i1-i6, що пояснюють рішення системи штучного інтелекту

На даному рисунку представлено правила у форматі пар індексів слів у вхідному повідомленні, а також наведено сумарну вагу правил для кожного з 6 варіантів пояснення. Відповідність індексів і слів представлено під списком правил. Пояснення i1 містить повний набір правил і всі ключові слова з індексами 0,1,16,18. Пояснення i6 містить лише одно правило (1,16) і складається з двох слів, для яких вказано порядок у вхідному повідомленні: Free ... deals. Відповідно, вага пояснень відрізняється в залежності від упорядкованої кількості слів-маркерів.

Порівняння пояснень з результатами фільтрації спаму на тестовому наборі показало відповідність пояснення і рішення системи.

## 7. Обговорення результатів дослідження

Розроблені методи побудови пояснень дають можливість сформулювати пояснення на різних рівнях представлення в залежності від доступності даних щодо процесу роботи системи ШІ та потреб користувача. Розроблені методи мають дві ключові відмінності. По-перше, враховується темпоральний аспект для процесу прийняття рішення, що дає можливість виявити залежності між окремими діями, які не відповідають обмеженням процесу прийняття рішень. Урахування темпорального аспекту також дає можливість врахувати вплив окремих

дій на результат роботи системи ШІ. По-друге, враховується упорядкованість вхідних даних при побудові пояснення, що дає можливість представити упорядковану підмножину вхідних даних як причину рішення системи. По-третє, використання підмножини правил дає можливість відбрати найточніші або найпростіші пояснення.

Обмеження на використання розроблених методів виникають внаслідок необхідності отримати записи виконання великої кількості імплементацій процесу прийняття рішення для розрахунку необхідних та достатніх умов істинності каузальних правил.

## **8. Висновки**

Виконано структурування представлення пояснення у темпоральному на каузальному аспектах на локальному, проміжному та глобальному рівнях. На локальному рівні пояснення представляються переважно у вигляді темпоральних залежностей між окремими діями процесу прийняття рішення. На проміжному рівні пояснення представляються у вигляді як темпоральних, так і каузальних залежностей, що характеризують процес прийняття рішення. На глобальному рівні пояснення представляються у вигляді каузальних залежностей, що визначають причини отриманого рішення.

Запропоновано темпорально-каузальний метод побудови пояснень, що містить етапи побудови темпоральних залежностей та каузальних обмежень щодо пар дій процесу прийняття рішення на локальному рівні деталізації пояснень, побудови темпоральних правил та каузальних залежностей щодо впливу дій процесу прийняття рішення на отриманий результат на проміжному рівні деталізації пояснень, побудови каузальних залежностей щодо впливу вхідних змінних на отриманий результат на глобальному рівні представлення пояснень.

Метод дає можливість на основі темпоральних залежностей сформувати пояснення у формі каузальних залежностей, що визначають дії процесу як причини отриманого рішення для внутрішніх користувачів та значення вхідних змінних як причини рішення для зовнішніх користувачів, що створює умови для поліпшення процесу прийняття рішення та відповідного підвищення рівня довіри користувачів.

Розроблено метод побудови пояснень на глобальному рівні представлення на основі упорядкованості вхідних даних. Метод враховує темпоральний та каузальний аспекти пояснення та містить етапи побудови темпоральних правил, визначення ваг правил на основі ваг антецедента та консеквента, побудови каузальних правил, побудови пояснення як сукупності зважених каузальних правил. Метод дає можливість пояснити причини отриманого в інтелектуальній системі рішення на основі інформації як про значення, так і про структуру вхідних даних, що створює умови для спрощення розуміння процесу прийняття рішення зовнішнім користувачем.

### **Перелік посилань:**

1. Bansal, G., Wu, T., Zhou, J., Fok, R., Nushi, B., Kamar, E., Ribeiro, M. T., & Weld, D. S. (2021). Does the whole exceed its parts? The effect of AI explanations on complementary team performance. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1-16. doi: <https://doi.org/10.1145/3411764.3445717>
2. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6, 52138-52160. <https://doi.org/10.1109/ACCESS.2018.2870052>
3. Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267, 1-38. doi: <https://doi.org/10.1016/j.artint.2018.07.007>
3. Byrne, R. M. (2019). Counterfactuals in explainable artificial intelligence (XAI): Evidence from human reasoning. *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19)*, 6276-6282. doi: <https://doi.org/10.24963/ijcai.2019/876>
4. Chalyi, S., Leshchynskyi, V., & Leshchynska, I. (2019). Method of forming recommendations using temporal constraints in a situation of cyclic cold start of the recommender system. *EUREKA: Physics and Engineering*, 4, 34-



40. doi: <https://doi.org/10.21303/2461-4262.2019.00952>

5. Chala O. (2018) Models of temporal dependencies for a probabilistic knowledge base. *Econtechmod. An International Quarterly Journal*, 7, 3, 53 – 58.

6. Chalyi, S., & Leshchynskyi, V. (2020). Temporal representation of causality in the construction of explanations in intelligent systems. *Advanced Information Systems*, 4(3), 113-117. doi: <https://doi.org/10.20998/2522-9052.2020.3.16>

7. Hoffman, R. R., Klein, G., & Miller, J. M. (2011). Naturalistic investigations and models of reasoning about complex indeterminate causation. *Information Knowledge Systems Management*, 10(1-4), 397-425. doi: <https://doi.org/10.3233/IKS-2012-0203>

8. D. Gunning i D. Aha, (2019). DARPA's Explainable Artificial Intelligence (XAI) Program. *AI Magazine*, 40(2), 44-58. doi: <https://doi.org/10.1609/aimag.v40i2.2850>

9. Hanif, A. et al. (2023). A Comprehensive Survey of Explainable Artificial Intelligence (XAI) Methods: Exploring Transparency and Interpretability. *Web Information Systems Engineering – WISE 2023. WISE 2023. Lecture Notes in Computer Science*, 14306. Springer, Singapore. doi: [https://doi.org/10.1007/978-981-99-7254-8\\_71](https://doi.org/10.1007/978-981-99-7254-8_71)

10. Yang, W., Wei, Y., Wei, H. et al. (2023). Survey on Explainable AI: From Approaches, Limitations and Applications Aspects. *Hum-Cent Intell Syst*, 3, 161–188. doi: <https://doi.org/10.1007/s44230-023-00038-y>

11. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). «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*, 1135-1144. doi: <https://doi.org/10.1145/2939672.2939778>

12. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In *Advances in neural information processing systems*. 4765-4774.

13. Чалий С. Ф., Лещинський В. О., Лещинська І. О. (2021). Контрфактуальна темпоральна модель причинно-наслідкових зв'язків для побудови поясень в інтелектуальних системах. *Вісник Національного технічного університету «ХПІ»*. Сер. : Системний аналіз, управління та інформаційні технології = *Bulletin of the National Technical University «KhPI»*. Ser.: System analysis, control and information technology : зб. наук. пр. Харків: НТУ «ХПІ», 2 (6), 41-46.

13. Chalyi, S., & Leshchynskyi, V. (2020). Causal reasoning in the construction of explanations in intelligent systems. *Advanced Information Systems*, 4(4), 105-110. doi: <https://doi.org/10.20998/2522-9052.2020.4.14>

14. Chala O. (2018). Models of temporal dependencies for a probabilistic knowledge base. *Econtechmod. An International Quarterly Journal*, 7, 3, 53–58.

15. Levykin V., Chala O. (2018). Development of a method of probabilistic inference of sequences of business process activities to support business process management. *Eastern-European Journal of Enterprise Technologies*, 5/3(95), 16-24. doi: [10.15587/1729-4061.2018.142664](https://doi.org/10.15587/1729-4061.2018.142664)

16. Чала О. В. (2020) Модель узагальненого представлення темпоральних знань для задач підтримки управлінських рішень. *Вісник Національного технічного університету «ХПІ»*. Системний аналіз, управління та інформаційні технології. № 1(3). С. 14-18. doi: [10.20998/2079-0023.2020.01.03](https://doi.org/10.20998/2079-0023.2020.01.03).

17. Chalyi, S., & Leshchynskyi, V. (2020). Causal reasoning in the construction of explanations in intelligent systems. *Advanced Information Systems*, 4(4), 105-110. doi: <https://doi.org/10.20998/2522-9052.2020.4.14>

18. Chalyi, S., & Leshchynskyi, V. (2021). Hierarchical representation of causal relationships to detail explanations in intelligent systems. *Advanced Information Systems*, 5(4), 103-108. doi: <https://doi.org/10.20998/2522-9052.2021.4.14>

19. Byrne, R. M. (2005). *The rational imagination: How people create alternatives to reality*. MIT press. doi: <https://doi.org/10.7551/mitpress/5756.001.0001>

20. Sloman, S. A. (2005). *Causal models: How people think about the world and its alternatives*. Oxford University Press. doi: <https://doi.org/10.1093/acprof:oso/9780195183115.001.0001>

Надійшла до редколегії 15.08.2024 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: 0000-0002-9982-9091

**Лещинський Володимир Олександрович**, кандидат технічних наук, доцент, доцент кафедри ПІ ХНУРЕ, м. Харків, Україна, e-mail: [volodymyr.leshchynskyi@nure.ua](mailto:volodymyr.leshchynskyi@nure.ua); ORCID: 0000-0002-8690-5702