

УДК 004.8:004.9

DOI: [https://doi.org/10.30837/bi.2026.1\(104\).11](https://doi.org/10.30837/bi.2026.1(104).11)С. Ф. Чалий¹, Р. В. Кравченко¹¹ХНУРЕ, м. Харків, Україна, serhii.chalyi@nure.ua,

ORCID iD: 0000-0002-9982-9091

¹ХНУРЕ, м. Харків, Україна, rostyslav.kravchenko1@nure.ua;

ORCID: 0009-0009-0324-3597

МЕТОД ПОБУДОВИ ТЕМПОРАЛЬНО УЗГОДЖЕНИХ ПОЯСНЕНЬ В ІНТЕЛЕКТУАЛЬНИХ СИСТЕМАХ НА ОСНОВІ ТЕМПОРАЛЬНИХ ГРАФОВИХ НЕЙРОННИХ МЕРЕЖ

Розглянуто проблему побудови пояснень з урахуванням порядку у часі вхідних даних інтелектуальних систем, вирішення якої потребує забезпечення темпоральної узгодженості каузальних зв'язків між подіями у згенерованих поясненнях. Розроблено інтервальний підхід до формалізації темпоральних відношень між подіями на основі тринадцяти базових відношень інтервальної алгебри з транзитивним поширенням обмежень для виявлення неявних темпоральних протиріч. Запропоновано метод побудови темпорально узгоджених пояснень, що включає обчислення внеску подій через інтегровані градієнти, гібридну генерацію пояснень природною мовою на основі шаблонів та донавченої великої мовної моделі, автоматизовану верифікацію темпоральної коректності з використанням інтервальної алгебри та оновлення фрагментів пояснень із темпоральними порушеннями. Метод забезпечує можливість формування пояснень для систем з упорядкованими у часі вхідними даними в критичних предметних областях, включаючи рекомендаційні системи електронної комерції, медичну діагностику, фінансові системи підтримки рішень, де необхідні як інтерпретованість рішень, так і верифікація коректності темпоральних залежностей між подіями.

ТЕМПОРАЛЬНО УЗГОДЖЕНІ ПОЯСНЕННЯ, ПОЯСНЮВАНИЙ ШТУЧНИЙ ІНТЕЛЕКТ, ТЕМПОРАЛЬНІ ГРАФОВІ НЕЙРОННІ МЕРЕЖІ, ІНТЕРВАЛЬНА АЛГЕБРА, ІНТЕГРОВАНІ ГРАДІЄНТИ, ВЕЛИКІ МОВНІ МОДЕЛІ

S.F. Chalyi, R.V. Kravchenko. Method for constructing temporally consistent explanations in intelligent systems based on temporal graph neural networks. The paper addresses the problem of constructing explanations that consider the temporal order of input data in intelligent systems, requiring the temporal consistency of causal relationships in the generated explanations. An interval-based approach for formalizing temporal relationships between events is developed based on thirteen basic relations of interval algebra with transitive constraint propagation to detect implicit temporal contradictions. A method for constructing temporally consistent explanations is proposed, which includes computing event contributions using integrated gradients, hybrid generation of natural-language explanations based on templates and a fine-tuned large language model, automated verification of temporal correctness using interval algebra, and updating explanation fragments with temporal violations. The method enables the generation of explanations for systems with time-ordered input data in critical domains, including e-commerce recommendation systems, medical diagnosis, and financial decision support systems, where both interpretability of decisions and verification of temporal dependency correctness are essential.

TEMPORALLY CONSISTENT EXPLANATIONS, EXPLAINABLE ARTIFICIAL INTELLIGENCE, TEMPORAL GRAPH NEURAL NETWORKS, INTERVAL ALGEBRA, INTEGRATED GRADIENTS, LARGE LANGUAGE MODELS

Вступ

Інтелектуальні системи на основі глибоких нейронних мереж досягають високої точності прогнозування поведінки користувачів у широкому спектрі предметних областей, включаючи рекомендаційні системи на платформах електронної комерції, медичну діагностику, фінансовий менеджмент. Проте такі системи з точки зору користувача функціонують як непрозора обчислювальна структура, що не відображає причинно-наслідкові зв'язки між вхідними даними та прогнозами системи. Відсутність пояснень прийнятих рішень зменшує довіру користувачів та знижує можливості впровадження пояснень для критичних застосувань, де помилкові прогнози можуть призвести до фінансових втрат, загроз безпеці чи порушень етичних норм.

Забезпечення пояснень для рішень інтелектуальних систем є нормативною вимогою згідно

з Регламентом Європейського Союзу щодо штучного інтелекту, у відповідності до якого мають бути надані зрозумілі пояснення логіки автоматизованого прийняття рішень для систем високого ризику, включаючи рекомендаційні системи електронної комерції та фінансові застосування [1]. Проте існуючі методи побудови пояснень для систем з упорядкованими у часі вхідними даними мають суттєвий недолік, пов'язаний з тим, що описана у поясненні каузальна послідовність подій суперечить фактичним міткам часу виникнення цих подій у вхідних даних системи. Зокрема, великі мовні моделі генерують значну кількість темпоральних невідповідностей у поясненнях звичайною мовою для послідовностей подій через відсутність явної перевірки міток часу.

Таким чином, проблема побудови пояснень з урахуванням порядку у часі у вхідних даних є актуальною.

Побудови пояснень виконуються в рамках наукового напрямку пояснювального штучного інтелекту (ХАІ). Даний напрямок досліджень спрямований на забезпечення зрозумілості рішень систем штучного інтелекту. Для побудови пояснень з урахуванням темпоральної упорядкованості вхідних даних в рамках даного дослідження розглядаються чотири групи методів. Графові методи обчислюють внески вузлів графової нейронної мережі (GNN) шляхом інтегрування градієнтів, виявляючи структурні патерни. Темпоральні графові методи розширюють статичні підходи механізмами пам'яті та агрегацією з урахуванням темпоральної еволюції. Методи на основі мовних моделей трансформують структуровані атрибути пояснень у пояснення природною мовою з використанням коротких структурованих промптів або гібридної генерації на основі патернів. Методи формалізації послідовностей подій у часі використовують темпоральні правила та інтервальну алгебру для представлення каузальних залежностей з темпоральними обмеженнями, що дає можливість сформувати патерни для генерації пояснень з використанням мовних моделей.

В роботі [2] показано, що градієнтні підходи для GNN забезпечують локальні пояснення, проте не генерують пояснення природною мовою та не верифікують коректність пояснень з урахуванням послідовності вхідних даних у часі. В дослідженні [3] запропонували інструментарій GNNExplainer для зменшення невідповідності між прогнозом та структурами підграфів, проте даний інструментарій не враховує послідовність вхідних подій у часі. В [4] розробили метод інтегрованих градієнтів, який задовольняє аксіоми повноти та чутливості, але не враховує темпоральну структуру вхідних даних. В роботі [5] запропонували Темпоральні графові нейронні мережі (TGNN) з модулем пам'яті для історичних станів та агрегацією з урахуванням часу, проте без механізмів пояснення з верифікацією. Дослідження [6] присвячено розробці TGNN з синусоїдальним кодуванням та багатоголовою увагою для визначення релевантних подій. Робота [7] пропонує метод адаптивного вибору інтервалів на основі відстані редагування графів та спектрального аналізу для детекції структурного дрейфу. В дослідженні [8] запропонували модель GNN з LSTM для виявлення темпоральних патернів та механізм уваги, що комбінує патерни окремих та взаємопов'язаних вершин.

Великі мовні моделі генерують пояснення на основі коротких промптів, однак створюють темпоральні невідповідності. В дослідженні [9] виявили темпоральні зміщення у LLM, тобто асиметрію точності для відношень «before» vs «after». Гібридні підходи комбінують шаблони з донавченими моделями. В [10] Wiegrefe et al. показали, що гібридна генерація підвищує контрольованість та фактуальну коректність.

Дослідження [11] присвячено розробці методу представлення знань на основі темпоральних правил виду «якщо подія А, то через деякий час Т відбудеться подія В, що забезпечує інтерпретабельність пояснення.

Інтервальна алгебра [12] на основі 13 базових відношень між інтервалами (before, after, meets, overlaps, during, starts, finishes, equals) дає можливість формалізувати темпоральні відношення. Алгебра забезпечує поширення обмежень на основі властивості транзитивності.

В дослідженні [13] застосували інтервальну алгебру для темпоральних баз даних. Робота [14] Leucker та Schallhart присвячена вирішенню проблеми верифікації в режимі реального часу з використанням лійної темпоральної логіки.

Проведений аналіз показав, що існуючі методи побудови пояснень не забезпечують автоматизовану верифікацію темпоральної узгодженості каузальних зв'язків між подіями. Методи на основі великих мовних моделей генерують текстові пояснення, однак не перевіряють мітки часу і тому генерують значну кількість темпоральних невідповідностей. Методи обчислення внеску вузлів нейронної мережі через градієнти обчислюють показники внеску вузлів, проте не перетворюють їх на пояснення природною мовою з верифікацією послідовності у часі на основі темпоральної логіки. Інтервальна алгебра забезпечує формальне представлення темпоральних відношень, проте не інтегрована в процес генерації пояснень для інтелектуальних систем на основі нейронних мереж. Методи представлення знань через темпоральні правила забезпечують інтерпретовану форму каузальних зв'язків, проте мають обмеження щодо масштабованості для великих графових структур.

Таким чином, необхідною є розробка методу, який інтегрує визначення впливу упорядкованих у часі вхідних подій, гібридну генерацію природною мовою на основі комбінації темпоральних шаблонів та тексту, згенерованого мовною моделлю, верифікацію темпоральних невідповідностей з використанням інтервальної алгебри у єдиний автоматизований процес генерації темпорально узгоджених пояснень.

1. Постановка задачі

Метою дослідження є розробка методу побудови темпорально узгоджених пояснень в інтелектуальних системах з автоматизованою верифікацією темпоральної коректності каузальних зв'язків між подіями, що відображають поведінку користувачів, та генерацією пояснень природною мовою.

Для досягнення мети дослідження вирішуються такі задачі: розробити інтервальний підхід до формування та верифікації темпорально узгоджених пояснень; розробити метод побудови темпорально узгоджених пояснень.

2. Інтервальний підхід до формування та верифікації темпорально узгоджених пояснень

Розроблений підхід передбачає формалізацію темпоральних відношень з використанням інтервальної алгебри, обчислення внеску подій у пояснення через інтегровані градієнти, генерацію пояснень природною мовою з донавчанням моделі, верифікацію темпоральної узгодженості для пояснення.

Базовий опис складових розробленого підходу наведено в табл. 1.

Таблиця 1

Ключові складові інтервального підходу

Компонент	Коментар	Призначення
1. Представлення темпоральних відношень (інтервальна алгебра)	Система правил для подій у часі одна відносно одної. Наприклад: «подія А ПЕРЕКРИВАЄ подію Б»	Забезпечити однозначне розуміння послідовності подій
2. Показник внеску події (інтегровані градієнти)	Числова оцінка того, наскільки конкретна подія вплинула на фінальне рішення системи. Показник може бути позитивним (подія збільшила ймовірність рекомендації) або негативним (зменшила)	Виявити, які саме події з історії користувача є найбільш важливими для пояснення. Наприклад, для системи e-commerce причиною рекомендації може бути минула покупка, перегляд тощо
3. Функція втрат для донавчання мовної моделі	Міра того, наскільки добре система навчилася генерувати пояснення у стилі та з термінами конкретної галузі	Забезпечити формування пояснень з використанням термінології предметної області
4. Верифікація темпоральної узгодженості	Автоматична перевірка того, чи не містить пояснення логічних помилок у часі. Наприклад, чи не каже система "А призвело до Б", коли насправді Б відбулося раніше А.	Виявити та виправити протиріччя в часі у поясненні

Формалізація темпоральних відношень між інтервалами часу подій здійснюється через інтервальну алгебру Аллена. Інтервал часу події представляється через дві часові мітки – момент початку події та момент завершення події. Алгебра визначає тринадцять базових типів відношень між двома інтервалами, які повністю описують усі можливі варіанти взаємного розташування інтервалів на часовій осі.

Базові відношення включають: before (до), after (після), meets (межує з), met-by (межується з), overlaps (перекриває), overlapped-by (перекривається), during

(всередині), contains (містить), starts (розпочинає), started-by (розпочинається одночасно з), finishes (завершує), finished-by (завершується одночасно з), equals (дорівнює).

Транзитивне поширення обмежень виконується через правила транзитивності інтервальної алгебри, які дозволяють автоматично виводити нові темпоральні відношення з відомих відношень. Наприклад, якщо подія А відбулася before події В, та подія В відбулася before події С, то автоматично виводиться, що подія А відбулася before події С, навіть якщо це відношення не згадано явно.

Показник внеску події обчислюється через метод інтегрованих градієнтів, який визначає вплив події на прогноз моделі.

Метод інтегрованих градієнтів обчислює внесок події у прогноз системи через вимірювання різниці між фактичним станом події h_t та нейтральним станом h_t^{base} , помноженої на середнє значення чутливості прогнозу до змін стану події вздовж шляху від нейтрального до фактичного стану. Це дозволяє визначити, наскільки конкретна подія збільшила або зменшила ймовірність прогнозу порівняно з ситуацією, коли ця подія була б відсутня або нейтральна. Відповідно, сума внесків усіх подій дорівнює різниці прогнозів. Тобто весь «ефект прогнозу» щодо системи повністю розподіляється між подіями без втрат. Такий підхід забезпечує інтерпретованість результатів і користувач може бути впевнений, що аналіз охоплює всі фактори, які вплинули на прогноз.

Функція втрат для донавчання мовної моделі обчислюється як:

$$\mathcal{L}_{gen} = -\frac{1}{L} \sum_{t=1}^L \log P(y_t | y_{<t}, S), \tag{1}$$

де L – довжина тексту у токенах; y_t – токен на позиції t ; S – структуроване пояснення з подіями та показниками внеску.

Функція втрат для донавчання вимірює, наскільки ймовірним є згенерований текст пояснення за умови структурованої інформації про події та їх внески. Мінімізація цієї функції змушує мовну модель навчитися генерувати пояснення природною мовою, які максимально відповідають структурованим даним та стилістичним нормам предметної області. Чим менше значення функції втрат, тим краще модель передбачає наступне слово у поясненні на основі попередніх слів та структурованої інформації.

При верифікації темпоральної узгодженості визначається показник TCS , який визначає частку коректних темпоральних відношень між подіями у згенерованому поясненні. Для кожної пари подій порівнюється темпоральне відношення, яке впливає з природномовного тексту пояснення, з фактичним темпоральним відношенням, обчисленим на основі міток часу з бази даних системи.

$$TCS = \frac{1}{K(K-1)} \sum_{i=1}^K \sum_{j \neq i}^K \mathbb{I}(R_{ij} = R'_{ij}), \quad (2)$$

де K – кількість подій; R_{ij} – фактичне темпоральне відношення; R'_{ij} – відношення з пояснення; $\mathbb{I}(\cdot)$ – індикаторна функція, яка приймає значення 1 за умови істинності виразу $R_{ij} = R'_{ij}$ та 0 в іншому випадку.

Значення $TCS = 1$ означає повну темпоральну узгодженість – усі темпоральні відношення у поясненні відповідають фактичним міткам часу; $TCS = 0$ означає повну невідповідність.

3. Метод побудови темпорально узгоджених пояснень

Розроблений метод формує темпорально узгоджені пояснення на основі інтеграції внеску подій, гібридної природномовної генерації пояснень та автоматизованої верифікації з використанням інтервальної алгебри. Метод виконується для кожного рішення інтелектуальної системи, що потребує пояснення.

Етап 1. Формування структурованого пояснення через обчислення внеску подій.

Вхідна послідовність подій з часовими мітками подається на темпоральну графову нейронну мережу для формування прогнозу.

Темпоральна графова нейронна мережа обробляє послідовність подій через механізм агрегації сусідів з урахуванням часових міток взаємодій та обчислює приховані репрезентації для кожної події.

Для кожної події обчислюється показник внеску в рішення, який показує, наскільки конкретна подія збільшила або зменшила ймовірність прогнозу порівняно з ситуацією відсутності цієї події. Обчислення виконується через інтегрування градієнтів функції прогнозу вздовж прямолінійного шляху від базової до фактичної репрезентації події, що забезпечує стабільність результатів відносно шуму у градієнтах.

При імплементації етапу формується структуроване пояснення з набору подій з найвищими абсолютними показниками внеску, впорядкованих за зростанням значень часу у часових мітках, з тим, щоб забезпечити коректну темпоральну послідовність.

Етап 2. Гібридна генерація пояснення природною мовою.

Структуроване пояснення подається на гібридний генератор, який комбінує попередньо визначений шаблон з явною темпоральною послідовністю та подає результат комбінування на донавчену велику мовну модель.

Шаблон вибирається з бібліотеки шаблонів залежно від типу прогнозу та кількості подій, з тим, щоб забезпечити правильну темпоральну послідовність через явні темпоральні залежності між заповнювачами місця для опису подій.

Донавчена мовна модель заповнює місця для опису подій шаблону, генеруючи текстові сегменти з

описом кожної події, зазначенням її внеску у прогноз та доменно-специфічною термінологією. Якість генерації забезпечується мінімізацією функції втрат згідно з формулою, яка змушує модель генерувати пояснення, що максимально відповідають структурованим даним та стилістичним нормам предметної області.

В результаті виконання етапу генерується пояснення природною мовою, яке описує каузальні зв'язки між подіями з дотриманням темпоральної послідовності згідно з шаблоном.

Етап 3. Автоматизована верифікація темпоральної узгодженості подій у поясненні.

Згенероване пояснення аналізується для вилучення всіх згаданих темпоральних відношень між подіями через розпізнавання темпоральних зв'язок типу «після», «внаслідок», «що призвело до».

Для кожної пари подій обчислюється фактичне темпоральне відношення на основі міток часу з бази даних згідно з тринадцятьма базовими відношеннями інтервальної алгебри Аллена.

Виконується транзитивне поширення обмежень через застосування правил транзитивності інтервальної алгебри, що дозволяє виявити неявні темпоральні протиріччя, які не є очевидними при аналізі окремих пар подій. Обчислюється показник темпоральної узгодженості як частка коректних темпоральних відношень у згенерованому поясненні.

Виявляють темпоральні невідповідності через порівняння відношень з тексту та фактичних відношень, обчислених на основі порівняння міток часу. Кожна виявлена невідповідність фіксується із зазначенням проблемного сегмента пояснення.

Етап 4. Перебудова фрагментів пояснення з темпоральними порушеннями.

Для кожної виявленої темпоральної невідповідності виконується перебудова відповідного фрагмента пояснення. Мовна модель отримує доповнений промпт з додатковим обмеженням у вигляді фактичного темпорального відношення між подіями, яке повинно бути відображено у згенерованому тексті з дотриманням семантичної узгодженості згідно з функцією втрат.

Перебудова фрагментів з темпоральними порушеннями виконується ітеративно, тобто після регенерації фрагмента знову виконується верифікація темпоральної узгодженості для перевірки: чи усунена невідповідність. Процес продовжується до досягнення $TCS = 1$, що означає повну темпоральну узгодженість, або до досягнення максимальної кількості ітерацій.

У випадку недосягнення повної темпоральної узгодженості після максимальної кількості ітерацій виконується повна перебудова пояснення з використанням альтернативного шаблону з явними вказівками на часові мітки подій.

4. Експериментальна перевірка розробленого методу

Експериментальна перевірка виконана з використанням публічного набору даних Retailrocket Recommender System Dataset, який містить 2 756 101 подію користувачів за період травень—вересень 2015 року.

Для експериментів використано підвибірку з 127 453 транзакцій, що включає послідовності подій: перегляди товарів (view), додавання у кошик (addtocart), покупки (transaction). Кожна подія містить ідентифікатор користувача, ідентифікатор товару, тип події та мітку часу у форматі UNIX timestamp.

Виділено фрагмент набору даних, який розділено на навчальну вибірку (80%), валідаційну вибірку (10%) та тестову вибірку (10%) з дотриманням темпоральної послідовності.

Для генерації пояснень використовується відкрита мовна модель LLaMA-2-7B з донавчанням методом LoRA з квантизацією 4 біти (Low-Rank Adaptation – метод ефективного донавчання великих моделей через низькорангові матриці адаптації) на наборі з 100 пар структурованих пояснень та референсних природномовних пояснень. Вибір LLaMA-2-7B та донавчання через LoRA обумовлений можливістю локального розгортання на мобільній робочій станції та ефективністю навчання з обмеженими обчислювальними ресурсами.

Експериментальний стенд реалізовано на мобільній робочій станції HP ZBook Studio з графічним прискорювачем NVIDIA. Програмна реалізація виконана з використанням фреймворку PyTorch 2.1 для навчання нейронних мереж, бібліотеки PyTorch Geometric для обробки графових структур, бібліотеки Transformers від Hugging Face для роботи з мовними моделями, бібліотеки PEFT для донавчання через LoRA.

Для оцінювання якості пояснень використано метрики UTS, TCS, TEC. User Trust Score (UTS – показник довіри користувачів) обчислюється як середнє значення оцінок довіри за шкалою від 1 до 10, де 1 – повна недовіра, 10 – повна довіра. Дана оцінка була адаптована для виключення суб'єктивної складової. Замість опитування користувачів було виконано аналіз результатів пошукових запитів. Значення оцінки розраховувалося як нормована кількість відповідей пошукової системи для вибірки із випадково вибраних прогнозів системи. Temporal Coherence Score (TCS) – як відношення кількості коректних темпоральних відношень до загальної кількості відношень. Temporal Error Count (TEC) обчислюється як абсолютна кількість виявлених темпоральних невідповідностей у 100 згенерованих поясненнях.

Для порівняння використовувались такі базові методи.

Метод 1. Генерація пояснень через GPT з типовими промптами без донавчання та верифікації отриманих пояснень.

Метод 2. Використання інструменту GNNExplainer для обчислення показників важливості і подальшої трансформації у текст через статичні шаблони.

Метод 3. Використання адаптованої LLaMA-2-7B через LoRA без гібридної генерації з шаблонами та без верифікації.

Метод 4. Генерація пояснень на основі темпоральних правил, що відображають зв'язки у часі між послідовностями подій.

Результати експериментальної перевірки розробленого методу представлено у табл. 2.

Таблиця 2

Порівняльний аналіз результатів методів генерації пояснень з темпорально упорядкованими даними

Метод	UTS (1–10)	TCS (0–1)	TEC (на 100 пояснень)
1. GPT	6,1	0,58	67
2. GNNExplainer та шаблон	5,3	0,45	78
3. LLaMA-2 -7B	6,8	0,65	52
4. Темпоральні правила	7,2	0,84	24
Розроблений метод	8,3	0,92	6

Розроблений метод досяг значення UTS 8,3 порівняно з 6,1 для великої мовної моделі.

Результати аналізу ефективності складових методу представлено у табл. 3.

Таблиця 3

Порівняльний аналіз ефективності етапів методу

Варіант методу	UTS (1–10)	TCS (0–1)	TEC (на 100 пояснень)
Повний метод	8,3	0,92	6
Метод без верифікації темпоральної узгодженості подій у поясненні	7,1	0,70	48
Метод без етапу перебудови фрагментів з темпоральними порушеннями	6,9	0,73	42
Метод без гібридної генерації (тільки з використанням шаблонів)	5,8	0,88	18
Метод без інтегрованих градієнтів (прості градієнти)	7,9	0,90	8

З таблиці 3 видно, що невиконання верифікації призводить до зниження TCS з 0,92 до 0,70 (тобто майже на 25%), збільшення TEC з 6 до 48 (збільшення кількості помилок у 8 разів), що підтверджує важливість автоматизованої верифікації для забезпечення темпоральної узгодженості пояснень. Заміна інтегрованих градієнтів на прості градієнти призводить

до незначного зниження всіх метрик, що вказує на корисність інтегрованих градієнтів для стабільності обчислень показників внеску, проте не є критичним компонентом для темпоральної узгодженості.

Результати експериментальної оцінки обчислювальної складності розробленого методу на основі часу виконання методу наведено у табл. 4.

Таблиця 4

Витрати часу на виконання етапів методу

Етап методу	Час (CPU)	Час (GPU)
Формування структурованого пояснення	28,3 с	2,1 с
Гібридна генерація	3,2 с	0,8 с
Верифікація темпоральної узгодженості	4,1 с	0,5 с
Регенерація сегментів (в середньому 1,8 ітерацій)	5,7 с	1,4 с
Загальний час	41,3 с	4,8 с

З таблиці 4 видно, що використання графічного процесора з об'ємом пам'яті 8 ГБ знижує час розрахунку пояснення у 8 разів, до 4,8 секунди, що свідчить про придатність розробленого методу до практичного застосування в режимі роботи ніарлайн.

Найбільш обчислювально затратним етапом є формування структурованого пояснення через обчислення інтегрованих градієнтів для всіх подій у послідовності. Використання графічного прискорювача знижує час виконання цього етапу з 28,3 секунди до 2,1 секунди (прискорення більш ніж у 13 разів).

Таким чином, розроблений метод формує пояснення на основі узгодженості між тим, як інтелектуальна система пояснює свої рекомендації, та реальною послідовністю подій у часі. Наприклад, коли користувач спочатку додає товар у кошик (подія А, момент часу t_1), потім переглянув схожий товар (подія В, момент часу t_2), і нарешті здійснив покупку (подія С, момент часу t_3), то інтелектуальна система не має генерувати пояснення «Ви купили товар С, тому що переглянули товар В», що є логічним, але порушує порядок у часі, оскільки подія В відбулася після додавання товару у кошик (подія А), яку система проігнорувала.

Розроблений метод має ряд переваг за рахунок того, що використовує нейросимвольний підхід до побудови пояснень. Нейронна мережа забезпечує природність мови та адаптивність до контексту. Символьний компонент, заданий інтервальною алгеброю, забезпечує перевірку коректності послідовності дій у часі. Тобто система не може згенерувати пояснення виду «подія А спричинила подію В», якщо насправді подія В відбулася раніше А.

Процедура автоматизованого виправлення помилок, виконує ітеративне переформування пояснення до тих пір, поки всі темпоральні відношення не стануть узгодженими, на відміну від традиційних

підходів, де згенероване пояснення є фінальним результатом.

Метод забезпечує адаптацію до предметної області за рахунок донавчання мовної моделі на обмеженому наборі прикладів пояснень, що дає можливість використовувати в поясненнях терміни з предметної області. Наприклад, для e-commerce це будуть терміни «сезонний тренд», «promotional campaign», «категорія товарів» у відповідному контексті. Такий підхід підвищує довіру користувачів, оскільки пояснення формуються аналогічно тому, що напише фахівець e-commerce.

Кожне пояснення може бути доповнене списком перевірених темпоральних відношень та їх статусом (виконується/не виконується). Такий підхід має переваги при побудові пояснень в критичних предметних областях, пов'язаних із ризиками для здоров'я, фінансів тощо. Система повинна пояснити не лише те, що вона рекомендує, а й деталізувати, чому ці пояснення є коректними з використанням темпоральної логіки.

Розроблений метод має ряд обмежень. По-перше, обчислювальні витрати суттєво залежать від кількості упорядкованих у часі подій, оскільки вклад обчислюється для кожної події у вхідному наборі даних. Відповідно, витрати часу на CPU склали 42,5 секунди, що не дає можливості використати метод в онлайн-режимі. По-друге, якість текстових пояснень значною мірою залежить від навчальних даних, які використовуються для донавчання мовної моделі. Доновчання потребує декілька сотень прикладів, і тому може бути проблематичним для предметних областей з обмеженою кількістю експертних пояснень.

Також інтервальна алгебра використовує обмежену кількість темпоральних відношень, що дає можливість формалізувати стандартні сценарії, проте не дає можливість описати деякі специфічні патерни. Наприклад, при неточному визначенні міток у вхідних даних розподіленої системи відбувається часткове перекриття подій на декілька хвилин, що не дає можливості точно врахувати темпоральну упорядкованість.

Зазначені обмеження визначають напрямки подальших досліджень, пов'язані із розпаралелюванням обчислень внеску подій, використанням ймовірнісної темпоральної логіки тощо.

Висновки

Розроблено метод побудови темпорально узгоджених пояснень в інтелектуальних системах, який інтегрує обробку послідовностей подій користувачів з часовими мітками з використанням графових нейронних мереж та автоматизовану верифікацію каузальної коректності згенерованих пояснень з використанням інтервальної алгебри. Метод включає етапи обчислення внеску подій через інтегровані

градієнти, гібридну генерацію пояснень звичайною мовою на основі шаблонів та донавченої великої мовної моделі, автоматизовану верифікацію обмежень з використанням інтервальної алгебри, таргетовану регенерацію сегментів з темпоральними порушеннями. Розроблений метод створює умови для впровадження пояснювальних інтелектуальних систем з можливістю інтерпретації рішень у критичних застосуваннях, де коректність темпоральних залежностей має високе значення при прийнятті управлінських рішень користувачами.

Подальші дослідження доцільно спрямувати на розширення методу для підтримки темпоральних логік з темпоральними кванторами та обмеженнями на тривалість інтервалів між подіями.

Експериментальна валідація на наборі даних рекомендаційної системи електронної комерції підтвердила підвищення показника довіри користувачів до пояснень, збільшення темпоральної узгодженості пояснень та зниження кількості помилок щодо послідовності в часі.

Список літератури

- [1] Regulation (EU) 2024/1689 of the European Parliament and of the Council on Artificial Intelligence (AI Act) // Official Journal of the European Union. 2024. L 1689. P. 1–144. URL: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689>
- [2] Longo L., Goebel R., Lecue F., Kieseberg P., Holzinger A. Explainable Artificial Intelligence: Concepts, Applications, Research Challenges and Visions // Machine Learning and Knowledge Extraction. Springer, 2020. Vol. 1. P. 1–16. DOI: https://doi.org/10.1007/978-3-030-57321-8_1
- [3] Ying R., Bourgeois D., You J., Zitnik M., Leskovec J. GNNExplainer: Generating Explanations for Graph Neural Networks // Advances in Neural Information Processing Systems (NeurIPS). 2019. Vol. 32. P. 9240–9251. URL: <https://proceedings.neurips.cc/paper/2019/hash/d80b7040b773199015de6d3b4293c8ff-Abstract.html>
- [4] Sundararajan M., Taly A., Yan Q. Axiomatic Attribution for Deep Networks // Proceedings of the 34th International Conference on Machine Learning (ICML). 2017. Vol. 70. P. 3319–3328. DOI: <https://doi.org/10.48550/arXiv.1703.01365>
- [5] Rossi E., Chamberlain B., Frasca F., Eynard D., Monti F., Bronstein M. Temporal Graph Networks for Deep Learning on Dynamic Graphs // Proceedings of the 37th International Conference on Machine Learning (ICML). 2020. P. 8230–8240. DOI: <https://doi.org/10.48550/arXiv.2006.10637>
- [6] Xu K., Hu W., Leskovec J., Jegelka S. How Powerful are Graph Neural Networks? // Proceedings of the 7th International Conference on Learning Representations (ICLR). 2019. P. 1–17. DOI: <https://doi.org/10.48550/arXiv.1810.00826>
- [7] Чалий С. Ф., Кравченко Р. В. Метод адаптивного вибору інтервалів часу для побудови графів темпоральних графових нейронних мереж // Вісник Національного технічного університету "ХПІ". Серія: Системний аналіз, управління та інформаційні технології. 2025. № 2(14). С. 129–139. DOI: <https://doi.org/10.20998/2079-0023.2025.02.19>
- [8] Чалий С. Ф., Кравченко Р. В. Графова нейронна мережа для темпорально упорядкованих даних в задачі побудови пояснень в інтелектуальній системі // Автоматизовані системи управління та прилади автоматики. 2025. Вип. 185. С. 77–86. DOI: <https://doi.org/10.30837/0135-1710.2025.185.077>
- [9] Tan Q., Ng H. T., Bing L. Towards Benchmarking and Improving the Temporal Reasoning Capability of Large Language Models // Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (ACL). 2023. Vol. 1. P. 14820–14835. DOI: <https://doi.org/10.18653/v1/2023.acl-long.828>
- [10] Wiegreffe S., Marasović A. Teach Me to Explain: A Review of Datasets for Explainable NLP // Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS). 2021. P. 1–14. DOI: <https://doi.org/10.48550/arXiv.2102.12060>
- [11] Levykin V., Chala O. Method of Determining Weights of Temporal Rules in Markov Logic Network for Building Knowledge Base in Information Control Systems // EUREKA: Physics and Engineering. 2018. No. 5. P. 3–10. DOI: <https://doi.org/10.21303/2461-4262.2018.00713>
- [12] Allen J. F. Maintaining Knowledge about Temporal Intervals // Communications of the ACM. 1983. Vol. 26. No. 11. P. 832–843. DOI: <https://doi.org/10.1145/182.358434>
- [13] Chomicki J., Toman D. Temporal Databases // Handbook of Temporal Reasoning in Artificial Intelligence. Elsevier, 2005. P. 429–467. DOI: [https://doi.org/10.1016/S1574-6526\(05\)80015-6](https://doi.org/10.1016/S1574-6526(05)80015-6)
- [14] Leucker M., Schallhart C. A Brief Account of Runtime Verification // Journal of Logic and Algebraic Programming. 2009. Vol. 78. No. 5. P. 293–303. DOI: <https://doi.org/10.1016/j.jlap.2008.08.004>

Received (Надійшла) 08.02.2026

Accepted for publication (Прийнята до друку) 02.03.2026

Publication date (Дата публікації) 27.03.2026