КОМПЛЕКСНЕ ПРЕДСТАВЛЕННЯ ПОЯСНЕННЯ У ТЕМПОРАЛЬНОМУ, КАУЗАЛЬНОМУ ТА ЦІЛЬОВОМУ АСПЕКТАХ

***Чалий С.Ф.**, д.т.н., професор, кафедра ІУС, ХНУРЕ*
***Лещинський В.О.**, к.т.н., доцент, кафедра ПІ, ХНУРЕ*

Інтелектуальні інформаційні системи (ІС) широко використовуються для підтримки й обґрунтування рішень в сферах охорони здоров'я, фінансів, маркетингу, освіти. Такі системи формують рішення за результатами збору й аналізу великих обсягів даних з використанням методів машинного навчання. Проте складність алгоритмів машинного навчання робить функціонування ІІС непрозорим для користувачів і утруднює розуміння того, як саме система прийшла до того чи іншого рішення. Тому непрозорість ІС може знижувати довіру до ІС та обмежувати ефективне використання рішень, що отримані у таких системах. Для вирішення проблеми непрозорості ІС використовуються пояснення [1]. Кожне пояснення надає опис причин поточного рішення інтелектуальної системи у зрозумілому для користувача форматі. Пояснення допомагають користувачам зрозуміти логіку роботи інтелектуальної системи, розкриваючи причинно-наслідкові зв'язки між даними, що використовуються ІС, та її рішенням. Також за допомогою пояснень користувачі можуть оцінити можливості використання отриманого в ІС результату для роз’вязанння практичних задач, отримавши інформацію про умови прийняття рішення. Представлення логіки роботи інтелектуальної системи та можливість оцінки практичного застосування рішення для вирішення задач користувача підвищують довіру до інтелектуальної системи і створюють умови для більш ефективного використання результатів її роботи. Проте існуючі розробки орієнтовані на побудову пояснення з урахуванням окремих аспектів рішення – каузальних зв'язків між вхідними даними та результатом, відповідності результату потребам користувача тощо [2, 3]. Розробці комплексного представлення пояснення, яке давало б можливість як зрозуміти логіку роботи ІС, так і оцінити можливості ефективного використання рішення, не приділяється достатньо уваги. Зазначене свідчить про актуальність розробки представлення пояснення, яке розкривало б не лише послідовність прийняття рішення в ІС та перелік причин (зокрема, вхідних і проміжних даних) отриманого рішення, а й давало можливість користувачеві оцінити умови практичного застосування цього рішення з урахуванням знань про предметну область.

Для вирішення даної задачі розроблено трехаспектне представлення пояснення, що охоплює темпоральний, каузальний та цільовий аспекти. У темпоральному аспекті пояснення представляється як процес, що описує послідовність прийняття рішення в інтелектуальній системі. У каузальному аспекті відоражається набір каузальних залежностей, які пояснюють причини прийнятого рішення. У цільовому аспекті визначаються ключові причини прийнятого рішення, які є важливими для практичного застосування отриманих результатів. Структурну схема пояснення наведено на рисунку 1.

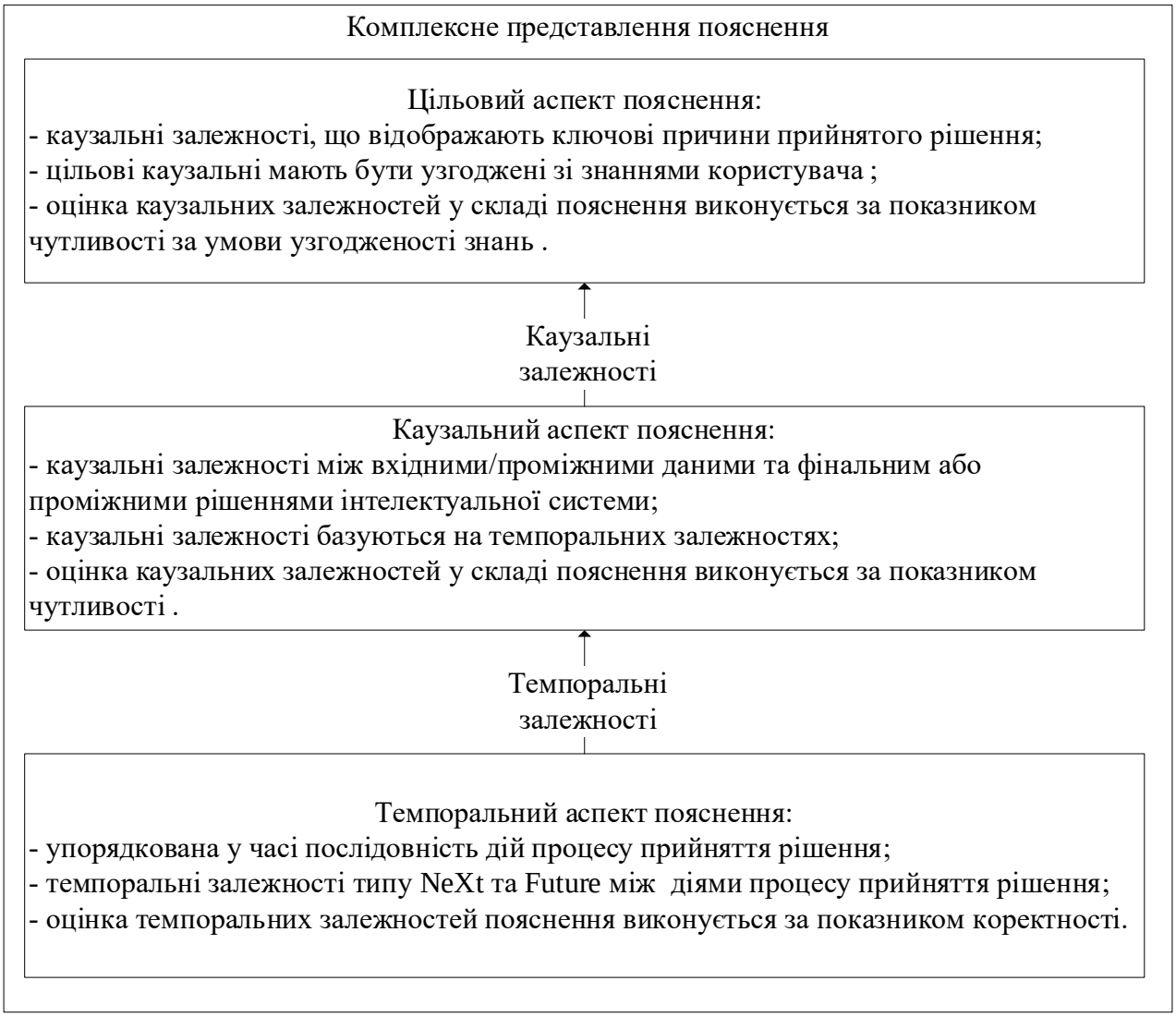

Рисунок 1 – Структура комплексного представлення пояснення

Формальне представлення пояснення містить множини темпоральних, каузальних і цільових залежностей $F_{i+n}^{i}, R_{i+n}^{i,j}, R_{j+n}^{i,V_k}$ відповідно: $P = \left\{ F_{i+n}^{i}, R_{i+n}^{i,j}, R_{j+n}^{i,V_k} \right\}$.

Запропонована узагальнене представлення пояснення надає комплексний опис послідовності прийняття рішення, визначає вхідні та проміжні дані (або дії), які є причинами рішення, а також містить ключові причини рішення з позицій його використання користувачем.

Список літератури

1. 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.

2. Chalyi, S. & Leshchynskyi, V. (2023). Інформаційна технологія оцінки пояснень в інтелектуальній інформаційній системі. Системи управління, навігації та зв'язку. Збірник наукових праць, (4), 120-124. 10.26906/SUNZ.2023.4.120.

3. Byrne RMJ. (2019). Counterfactuals in explainable artificial intelligence (XAI): evidence from human reasoning. Proceedings of the twenty-eighth international joint conference on artificial intelligence, IJCAI 2019, Macao, China, August 10–16, 2019, ijcai.org. (p 6276-6282).