

МОДЕЛЬ ВЕРИФІКАЦІЇ РІШЕНЬ ФІНАНСОВИХ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ

Бараней Д.І., Філімончук Т.В.

Харківський національний університет радіоелектроніки, Харків, Україна

На даний час фінансовий сектор активно використовує інтелектуальні системи для автоматизації кредитного аналізу, виявлення шахрайства, управління ризиками та підтримки прийняття рішень. Такі системи, побудовані на основі сучасних методів машинного навчання та глибинних нейронних мереж та дозволяють значно підвищити ефективність обробки великих обсягів даних, зменшити вплив людського фактору та скоротити час ухвалення рішень. Водночас зростання складності моделей призвело до формування так званої проблеми «чорної скриньки», що істотно ускладнює інтерпретацію, аудит та верифікацію прийнятих рішень [1]. Це особливо критично у фінансовій сфері, де будь-яке рішення повинно бути не лише точним, обґрунтованим та прозорим, а й таким, що відповідає регуляторним вимогам.

У наукових дослідженнях та практиці фінансових установ основна увага наразі зосереджена на підходах Explainable AI (xAI), які спрямовані на підвищення прозорості моделей шляхом пояснення отриманих результатів [2]. Зокрема, застосовуються такі методи, як локальні пояснення (LIME), адитивні моделі внесків ознак (SHAP), візуалізація важливості ознак, а також побудова сурогатних моделей. Вони дозволяють отримати уявлення про те, які фактори вплинули на конкретне рішення моделі.

Однак існуючі методи здебільшого забезпечують пояснення, а не повноцінну верифікацію рішень, що пов'язано з відсутністю єдиного підходу до визначення критеріїв пояснюваності, інтерпретованості та достовірності моделей.

У науковій спільноті відзначається розмитість цих понять та відсутність узгоджених метрик оцінювання, що ускладнює порівняння підходів та побудову універсальних моделей верифікації [2]. Крім того, пояснення, отримані за допомогою xAI-методів, можуть бути нестабільними або вводити в оману щодо реальної логіки роботи моделі, оскільки вони часто є лише апроксимацією її поведінки [3].

Це створює ризики при використанні таких методів у високоризикових фінансових застосуваннях, зокрема у кредитному скорингу, боротьбі з відмиванням коштів та оцінці інвестиційних ризиків.

Таким чином, існуючий стан розвитку характеризується переходом від простих інтерпретованих моделей до складних високоточних, але непрозорих систем, для яких класичні підходи пояснення є недостатніми. Це зумовлює необхідність розроблення нових підходів до верифікації рішень, які забезпечують не лише інтерпретацію, а й формальне підтвердження їх коректності, узгодженості та відповідності предметній області.

У цьому контексті перспективним напрямом є використання онтологічних моделей та нейросимвольних підходів, які дозволяють поєднати статистичні

методи машинного навчання з формалізованими знаннями предметної області. Подальші дослідження у цьому напрямі можуть бути спрямовані на розроблення гібридних архітектур, які забезпечують інтеграцію пояснюваності, верифікації та відповідності нормативним вимогам, а також на створення стандартизованих метрик оцінювання якості пояснень і достовірності рішень. Це сприятиме формуванню більш надійних, прозорих і відповідальних інтелектуальних систем у фінансовому секторі.

Метою доповіді є обґрунтування складових багаторівневої онтологічної моделі представлення знань про процеси формування автоматизованих аналітичних результатів.

Запропонований підхід орієнтований на інтеграцію різномірних джерел знань та забезпечення узгодженого опису як внутрішніх характеристик моделей штучного інтелекту, так і зовнішніх обмежень предметної області.

На відміну від існуючих рішень, розроблена модель передбачає спільне представлення метаданих інтерпретації моделей штучного інтелекту та нормативних вимог, що регламентують процес прийняття рішень, що дозволяє поєднати формалізовані описи алгоритмічних процедур із експертними знаннями, які зазвичай мають детермінований характер та задаються у вигляді правил, обмежень або категорій.

Важливою особливістю моделі є забезпечення узгодження між стохастичною природою параметрів моделей машинного навчання та детермінованими характеристиками предметної області. Завдяки цьому стає можливим формалізувати логічні зв'язки між результатами обчислювальних процедур та експертними інтерпретаціями, що підвищує прозорість та пояснюваність отриманих результатів.

Запропонований підхід створює цілісний методологічний базис для семантичної верифікації результатів інтелектуального аналізу даних, що сприяє підвищенню довіри до автоматизованих систем, забезпечує можливість контролю їх відповідності нормативним вимогам та відкриває перспективи для подальшого розвитку систем підтримки прийняття рішень на основі штучного інтелекту.

Список літератури

1. John I. Explainable AI Models for Financial Decision Systems. – 2024. – [Електронний ресурс]. – Режим доступу: https://www.researchgate.net/publication/393453395_Explainable_AI_Models_for_Financial_Decision_Systems (дата звернення: 10.04.2026).
2. Černevičienė J., Kabašinskas A. Explainable artificial intelligence (XAI) in finance: a systematic literature review. Artificial Intelligence Review. Vol. 57, article number 216, 2024. DOI: <https://doi.org/10.1007/s10462-024-10854-8>
3. Evite P.M., Svetlova E., Bucur D. Trade-offs in Financial AI: Explainability in a Trilemma with Accuracy and Compliance. arXiv preprint arXiv:2602.01368v1. 2025. DOI: <https://doi.org/10.48550/arXiv.2602.01368>