ISSN 3083-7715 (online) *Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188)*
ISSN 3083-7650 (0135-1710)(print) *Management Information System and Devices. 2026. No. 1 (188)*

УДК 004.8:004.9 DOI: https://doi.org/10.30837/0135-1710.2026.188.098

Чалий С. Ф., Лещинська І. О.

НЕЙРОСИМВОЛЬНИЙ ФРЕЙМВОРК ФОРМУВАННЯ МЕНТАЛЬНИХ МОДЕЛЕЙ РІШЕНЬ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ

Предметом роботи *є процес побудови ментальних моделей рішень інтелектуальних систем для користувачів з різним рівнем компетентності та відтворення послідовностей їх взаємодії з інтелектуальними системами в структуроване ментальне подання.* ***Мета дослідження*** *– розробити нейросимвольний підхід, що інтегрує концептуальну ментальну модель, систему принципів побудови й нейросимвольну реалізацію ментальної моделі для забезпечення персоналізованих пояснень в інтелектуальних системах.* ***Завдання:*** *розробити нейросимвольний фреймворк побудови ментальних моделей; розширити систему принципів для персоналізації пояснень; експериментально перевірити нейросимвольний фреймворк формування ментальних моделей рішень інтелектуальних систем.* ***Методи дослідження*** *основані на застосуванні нейросимвольного штучного інтелекту, що поєднує машинне навчання з причинно-наслідковим символьним міркуванням, використанні операторів лінійної темпоральної логіки для верифікації каузальних залежностей і адаптивному механізмі відбору концептів за рівнем компетентності користувача.* ***Результати дослідження.*** *Розроблено нейросимвольний фреймворк формування ментальних моделей рішень інтелектуальних систем, який містить: ментальну модель, що інтегрує множини позитивних і негативних властивостей рішення, каузальний граф причинно-наслідкових залежностей і лінгвістичні мітки для інтерпретації структури моделі природною мовою; запропоновано розширену систему принципів побудови ментальних моделей, яка до наявних принципів відповідності, множинності, неповноти й доповнення долучає нові принципи деталізації, темпоральної узгодженості та поведінкового виведення; створити адаптивну порогову функцію відбору концептів, що визначає рівень деталізації ментальної моделі відповідно до рівня компетентності конкретного користувача. Доведено, що система з чотирьох принципів є необхідною, але не достатньою умовою коректності нейросимвольної ментальної моделі. Експериментальна перевірка підтвердила суттєве підвищення темпоральної узгодженості каузальних залежностей порівняно з наявними методами.* ***Висновки.*** *Запропонований фреймворк забезпечує побудову персоналізованих ментальних моделей з верифікованими каузальними залежностями, відповідає вимогам прозорості моделей штучного інтелекту й забезпечує масштабовану персоналізацію пояснень для інтелектуальних систем.*

Ключові слова: *нейросимвольний штучний інтелект; ментальна модель; інтелектуальна інформаційна система; пояснювальний штучний інтелект; персоналізація; темпоральна узгодженість; каузальне міркування; формування пояснень.*

Вступ

Інтелектуальні інформаційні системи (ІІС) широко застосовуються для підтримки прийняття рішень у таких сферах, як медицина, фінанси, освіта, промисловість. Зростання складності внутрішніх моделей ІІС, зокрема внаслідок використання глибоких нейронних мереж, призводить до формування рішень такими системами в режимі непрозорої "чорної скриньки", що суттєво ускладнює інтерпретацію цих рішень кінцевими користувачами. Нормативні європейські вимоги EU AI Act [1] та GDPR [2] орієнтують розробників ІІС на забезпечення прозорості прийняття рішень за допомогою формування зрозумілих для користувачів пояснень. Формування пояснень досліджують у межах наукового напряму пояснювального штучного інтелекту (XAI, *Explainable Artificial Intelligence*). Традиційно методи XAI аналізують внесок вхідних ознак у рішення ІІС і на цій основі генерують узагальнені пояснення. Такі пояснення не беруть до уваги особливості сприйняття конкретного користувача, що спричиняє зниження рівня довіри до ІІС, а також ускладнює практичне застосування досягнутих результатів.

Для подолання вказаного обмеження й забезпечення персоналізованих пояснень можуть бути впроваджені ментальні моделі. Останні є внутрішнім когнітивним поданням, яке користувачі використовують для інтерпретації поведінки складних систем і прогнозування результатів їх роботи [3]. Тобто ментальна модель визначає розуміння користувачами ІІС причинно-наслідкових зв’язків у отриманому рішенні, тому є основою для формування адаптивних персоналізованих пояснень [4]. Побудова ментальних моделей

Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188) *ISSN 3083-7715 (online)*
Management Information System and Devices. 2026. No. 1 (188) *ISSN 3083-7650 (0135-1710)(print)*

на основі інформації щодо поведінки інтелектуальної системи є перспективним напрямом персоналізації XAI. Проте наявні підходи до побудови ментальних моделей орієнтовані насамперед на експертне виявлення когнітивних схем за результатами інтерв'ю [5].

Перспективним напрямом для розв'язання задачі автоматизованого формування ментальних моделей є використання нейросимвольного штучного інтелекту (NeSy, Neurosymbolic AI). Останній поєднує традиційне машинне навчання з причинно-наслідковим міркуванням у межах символьних систем, що створює умови для побудови інтерпретованих моделей рішень ІІС [6, 7]. Останнім часом розроблено низку складників NeSy-підходу до побудови ментальних моделей рішень ІІС, зокрема концептуальну модель, що описує структуру ментального подання рішення через знання, концепти й досвід користувача [8]; систему принципів, що визначають вимоги до побудови ментальних моделей зовнішніх користувачів [9]; двошарову NeSy-архітектуру, яку можна застосовувати в побудові ментальної моделі [10], а також метод розроблення цієї нейросимвольної архітектури [11]. Перелічені складники є елементами, на основі яких може бути побудовано єдину систему формування ментальної моделі рішення ІІС з використанням нейросимвольного підходу, яка має задовольняти вимоги щодо прозорості систем штучного інтелекту, визначених EU AI Act Article 13 [1].

Зазначене свідчить про актуальність розроблення комплексного нейросимвольного фреймворку, що інтегрує в єдину систему концептуальну ментальну модель, її нейросимвольну реалізацію, а також принципи побудови ментальних моделей з огляду на рівень компетентності користувачів на основі адаптивного механізму відбору концептів для побудови персоналізованих пояснень.

Аналіз літературних джерел і постановка проблеми дослідження

Концепцію ментальних моделей як когнітивного механізму інтерпретації складних систем запропоновано в роботі [3]. Відповідно до цієї концепції користувач формує внутрішній спрощений образ системи, з якою він взаємодіє. Цей спрощений образ дає змогу прогнозувати поведінку системи й оцінювати якість отриманих рішень. Тому ментальна модель безпосередньо визначає ступінь довіри до рекомендацій ІІС і ефективність застосування останніх для розв'язання практичних завдань користувача [4]. Парадигма пояснювального XAI 2.0 обґрунтована в дослідженні [12]. Як актуальний напрям розвитку пояснювального штучного інтелекту XAI 2.0 розглядає персоналізацію пояснень, що може бути реалізована на основі ментальних моделей.

Методи XAI першого покоління містять LIME та SHAP. Метод локальних агностичних пояснень (LIME, *Local Interpretable Model-Agnostic Explanations*) [13] апроксимує поведінку складного класифікатора через інтерпретовану локальну модель, що формується в межах досліджуваного прикладу. Пояснення не залежить від архітектури ІІС. Проте LIME генерує неперсоналізоване спільне для всіх користувачів пояснення без огляду на рівень підготовки останніх. Цей метод не бере до уваги темпоральний аспект причинно-наслідкових зв'язків, тому не призначений для побудови ментальної моделі рішення ІІС.

Метод Шеплі (SHAP, *SHapley Additive exPlanations*) [14] розроблений на основі теорії кооперативних ігор. Він розподіляє внесок кожної вхідної ознаки в отримане рішення. Метод SHAP реалізує глобальні та локальні пояснення. Однак він формує уніфіковані пояснення, ігноруючи особливості сприйняття цих пояснень користувачем. Це обмежує можливості його застосування для побудови ментальних моделей.

Зазначені методи призначені для побудови пояснень до рішень вже наявної ІІС. Можливості інтеграції функціональності традиційної архітектури ІІС з безпосередньою інтерпретацією рішень досягаються з упровадженням нейросимвольних методів.

Нейросимвольний штучний інтелект [6, 7] призначений для подолання обмеження традиційних нейронних мереж, пов'язаного з відсутністю символьного виведення для верифікації процесу досягнення результату. Поєднання традиційного машинного навчання з причинно-наслідковим міркуванням символьних систем створює умови для побудови моделей формування рішень, які можуть бути безпосередньо інтерпретовані.

Моделі концептного вузького місця (CBM, *Concept Bottleneck Models*) [15] інтегрують у архітектуру нейронної мережі явний концептний шар перед виходом класифікатора. Концепти забезпечують інтерпретацію зв'язку між ознаками й рішенням.

ISSN 3083-7715 (online) Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188)
ISSN 3083-7650 (0135-1710)(print) Management Information System and Devices. 2026. No. 1 (188)

Проте множина концептів задається розробником апріорно. Відповідно, персоналізація в межах цього підходу не забезпечується.

Логічні тензорні мережі (LTN, *Logic Tensor Networks*) [16] реалізують нейросимвольне виведення через тензорне подання логічних аксіом і предикатів. LTN дають змогу задавати обмеження у вигляді логічних формул і навчати мережу, зважаючи на ці обмеження. Однак LTN не підтримують персоналізацію та не беруть до уваги темпоральний аспект у причинно-наслідкових залежностях, що не забезпечує узгодженості знань у ментальній моделі.

Нейросимвольне ймовірнісне логічне програмування (*DeepProbLog*) [17] поєднує нейронні компоненти через імовірнісні предикати, що уможливлює навчання мережі для розв'язання логічних завдань. Проте персоналізація відповідно до рівня компетентності користувача й перевірка темпоральної узгодженості в межах цього підходу не підтримується.

Нейросимвольне навчання концептів [18] забезпечує виведення абстрактних концептів зі спостережень у неструктурованих сценах. Однак цей підхід спрямований на аналіз статичних сцен, що не дає змоги застосувати його для побудови ментальної моделі.

Принципи побудови ментальних моделей рішень ІС для зовнішнього користувача уточнено в роботі [9]. Традиційні принципи відповідності, коли модель структурно має відповідати поясненням рішень ІС, множинності, коли ментальні моделі різних користувачів є унікальними, неповноти, коли модель є когнітивно спрощеним поданням рішення, розширено принципом доповнення, згідно з яким модель має зважати на негативні властивості рішення ІІС. Проте питання персоналізації ментальних моделей у цьому дослідженні не розглянуто.

Отже, наявні принципи й підходи до побудови ментальних моделей, які забезпечують пояснюваність рішень інтелектуальних систем, не беруть до уваги можливості побудови моделей відповідно до рівня розуміння користувача, а також не забезпечують темпоральну верифікацію застосованих для побудови пояснень причинно-наслідкових залежностей.

Мета й завдання роботи

Метою дослідження є розроблення нейросимвольного підходу, що інтегрує концептуальну ментальну модель рішення інтелектуальної системи, принципи побудови ментальних моделей, а також нейросимвольну реалізацію ментальної моделі, щоб забезпечити можливість побудови персоналізованих пояснень в інтелектуальних системах.

Для досягнення окресленої мети необхідно виконати такі завдання: розробити нейросимвольний фреймворк побудови ментальних моделей; запропонувати принципи побудови ментальних моделей для персоналізації пояснень в інтелектуальних системах; експериментально перевірити нейросимвольний фреймворк формування ментальних моделей рішень інтелектуальних систем.

Нейросимвольний фреймворк формування персоналізованої ментальної моделі

Нейросимвольний фреймворк призначений для відтворення послідовностей взаємодії користувача з ІІС у ментальну модель рішення інтелектуальної системи:

$$\Omega: B_i \to M_i, \qquad (1)$$

де B_i – упорядкована темпоральна послідовність подій, що відображають взаємодію $i-$ користувача з ІІС; M_i – ментальна модель, що охоплює: множину V_i^+ позитивних властивостей рішення ІІС; множину V_i^- негативних властивостей рішення ІІС; граф G_i, що відтворює каузальні залежності між властивостями рішення; множину лінгвістичних міток L_i, призначених для інтерпретації структури моделі природною мовою.

Визначення 1

Нейросимвольною ментальною моделлю рішення інтелектуальної системи називається чотирикомпонентна структура M_i, яка містить множини позитивних і негативних властивостей рішення системи, каузальні зв'язки між властивостями рішення, а також лінгвістичні мітки для інтерпретації структури моделі. Модель формується на основі даних щодо послідовності взаємодії користувача з інтелектуальною інформаційною системою.

Персоналізація цієї моделі в межах розробленого фреймворку потребує розширення запропонованої в праці [9] системи принципами відповідності, множинності, неповноти й доповнення.

Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188) *ISSN 3083-7715 (online)*
Management Information System and Devices. 2026. No. 1 (188) *ISSN 3083-7650 (0135-1710)(print)*

Твердження

Система принципів відповідності, множинності, неповноти й доповнення є необхідною, але недостатньою умовою коректності нейросимвольної ментальної моделі рішення інтелектуальної системи в тому сенсі, що може існувати модель M_i, яка задовольняє ці чотири принципи, але порушує щонайменше одну з вимог: автоматизованого виведення ментальної моделі з послідовності взаємодії користувача з інтелектуальною системою (а); адаптації деталізації моделі до рівня компетентності користувача (б); темпоральної несуперечливості каузальних залежностей графа G_i (в).

Проведемо доказ від супротивного на основі трьох відповідних контрприкладів.

Контрприклад щодо вимоги (а). Нехай експерт предметної галузі побудував ментальну модель традиційно вручну, незважаючи на поведінкові дані користувача. Тоді вона задовольняє всі чотири принципи відповідності, множинності, неповноти й доповнення, проте не може бути відтворена й верифікована на основі інформації щодо взаємодії користувача з інтелектуальною системою, оскільки модель не містить механізму реалізації відображення (а).

Контрприклад щодо вимоги (б). Ментальна модель M_i, яка відповідає чотирьом принципам, може містити кількість концептів, що перевищує когнітивну ємність користувача-початківця, визначену законом Міллера: 7±2 унікальних елементи [19, 20]. Отже, ця модель не може бути використана з метою адаптації пояснення для такого користувача. Ця суперечність пов’язана з тим, що принципи відповідності, множинності, неповноти й доповнення визначають склад і структуру властивостей рішення, але не обмежують кількість ключових властивостей, доступних для сприйняття користувачем.

Контрприклад щодо вимоги (в). Нехай каузальний граф у складі ментальної моделі містить дугу, де наслідок передує причині в часі. Такий граф задовольняє розглянуті чотири принципи, оскільки останні визначають склад властивостей рішення, а не їх упорядкованість у часі. Тобто вимога (в) не є наслідком принципів відповідності, множинності, неповноти й доповнення.

Три розглянуті контрприклади є попарно незалежними, оскільки жоден з них не використовує умови іншого прикладу, що і свідчить про істинність цього твердження.

Отже, нейросимвольний фреймворк визначає реалізацію відображення (1) на основі принципів відповідності, множинності, неповноти й доповнення. Проте в побудові персоналізованих ментальних моделей ця система принципів потребує розширення.

Розширена систем принципів побудови ментальних моделей

У межах нейросимвольного фреймворку наявна система принципів побудови ментальних моделей має бути розширена з огляду на вимоги (а)–(в). Відповідно, розширена система містить принципи деталізації, темпоральної узгодженості та поведінкового виведення.

Принцип деталізації визначає, що ментальна модель повинна мати деталізацію відповідно до рівня підготовки користувача. Початківець зазвичай не здатний проаналізувати стільки ж концептів у складі рішення, що й фахівець. Надлишкова деталізація спричиняє перевантаження й пояснення стає "непрозорим" для користувача. Згідно з принципом деталізації, кількість активних концептів V_i^+ у ментальній моделі рішення має монотонно зростати з підвищенням компетентності користувача. Тобто

$$\Delta(M_i, u) = f(u) : \Delta_{\exp ert} > \Delta_{practic} > \Delta_{junor}, \qquad (2)$$

де $\Delta(M_i, u) \in [0,1]$ – рівень деталізації ментальної моделі M_i відповідно до рівня u підготовки користувача; $f(u)$ – монотонна зростаюча функція компетентності, яка має найбільше значення $\Delta_{\exp ert}$ для користувача-експерта й найменше значення Δ_{junor} для новачка.

Принцип темпоральної узгодженості визначає обов’язкову темпоральну упорядкованість каузального графа в ментальній моделі M_i, тобто причина завжди має передувати наслідку.

Важливість цього принципу пов’язана з тим, що побудований на основі аналізу взаємодії користувача з ІІС граф може містити дуги з некоректним поданням часу, які не мають каузальної інтерпретації. Тому цей принцип задає обмеження в побудові причинно-наслідкового графа. Вимога темпоральної узгодженості задається з використання операторів **G** та **F** лінійної темпоральної логіки (LTL). Оператор **G** (*globally*) вимагає виконання умови в усіх майбутніх станах

ISSN 3083-7715 (online) Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188)
ISSN 3083-7650 (0135-1710)(print) Management Information System and Devices. 2026. No. 1 (188)

системи. Оператор **F** (*future*) задає виконання умови колись у майбутньому стані. Тоді для довільної дуги (d_l, d_m) з графа G_i, для якої обов'язково буде істинним стан c_l, завжди в майбутньому має бути істинним стан d_m:

$$\forall(d_l, d_m) \in G_i : s(d_l, d_m) \Rightarrow \mathbf{G}d_l \rightarrow \mathbf{F}d_m. \quad (3)$$

Цей принцип дає змогу вилучити в процесі побудови графа G_i дуги (d_l, d_m), що не задовольняють умову (3).

Принцип поведінкового виведення передбачає, що персоналізована ментальна модель має зважати на послідовність взаємодії з ІІС конкретного користувача (а не лише представлення експерта щодо об'єктів предметної галузі). Тобто ментальна модель рішення має бути функцією поведінкових даних за умови достатнього розміру n набору цих даних:

$$M_i = \Omega(B_i) \big| |B_i| \geq n. \quad (4)$$

Запропонований принцип забезпечує відтворюваність ментальної моделі, що створює умови для побудови персоналізованих пояснень.

Відповідність розробленої системи принципів і описаної в дослідженні [10] нейросимвольної архітектури ментальної моделі подано в табл. 1.

Таблиця 1. *Реалізація принципів побудови ментальної моделі в межах нейросимвольної архітектури*

Принцип	Модуль архітектури	Механізм забезпечення
1) Відповідність структури	NeSy-транслятор	Перетворює вектор латентного простору на символьний концепт із онтології, тобто забезпечує зв'язок між нейромережним і символьним рівнями архітектури
2) Множинність	VAE-кодувач	Кодує індивідуальну послідовність взаємодії користувача з ІІС у латентний прихований простір
3) Неповнота	Механізм уваги	Фільтрує концепти за релевантністю до поточного контексту рішення, тобто в ментальну модель додаються лише елементи, для яких перевищено поріг значущості для певного типу користувача
4) Доповнення	Модуль негативних ознак	Формує негативний контур ментальної моделі, що передбачає вилучення властивостей, що обмежують використання рішення
5) Деталізація	Адаптивний класифікатор типу користувача	Налаштовує рівень деталізації ментальної моделі під конкретний тип користувача, тобто новачок отримує узагальнене подання рішення, а фахівець – деталізоване подання
6) Темпоральна узгодженість	LTL-верифікатор	Перевіряє відповідність послідовності концептів у ментальній моделі порядку використання (аналізу) рішення користувачем і вилучає інвертовані в часі каузальні ланцюжки
7) Поведінкове виведення	Модуль збору поведінкових даних користувача	Логування даних про взаємодію користувача із ІІС

Реалізація принципу адаптивної деталізації потребує формування механізму адаптації моделі. Такий механізм імплементовано через адаптивний поріг відбору концептів, що описують властивості рішення ІІС. Експерименти, які провели автори в попередніх дослідженнях, дали змогу визначати базовий поріг відбору на рівні $\theta^* = 0{,}85$. Однак такий фіксований поріг може спричинити перенавантаження новачків. З іншого боку, набір понять у ментальній моделі може бути недостатнім для фахівців.

Для побудови адаптивного порогу пропонується лінійна модель уточнення значення порогу θ залежно від $u-$ рівня підготовки користувачів:

$$\theta(u) = \theta^* - k \cdot \Delta(M_i, u), \quad (5)$$

де k – коефіцієнт адаптації.

Семантика формули (5) полягає в такому: вищий рівень компетентності означає збільшення $\Delta(M_i, u)$, що приводить до зниження $\theta(u)$ за умови заданого базового рівня адаптації θ^*. Залежність є монотонно спадною в разі збільшення $\Delta(M_i, u)$, що задовольняє умову (2).

Експериментальна перевірка

Експериментальну перевірку розробленого фреймворку виконано із застосуванням набору відгуків щодо ноутбуків на українській маркетплейс-платформі. Відповіді щодо ноутбуків подано в текстовій формі. Тобто ідея експерименту полягає в тому, щоб на підставі аналізу відгуків оцінити рекомендації щодо товарів і послуг у системі електронної комерції.

На основі текстових відповідей сформовано датасет із відгуків користувачів. Особливість

Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188) *ISSN 3083-7715 (online)*
Management Information System and Devices. 2026. No. 1 (188) *ISSN 3083-7650 (0135-1710)(print)*

датасету полягає в тому, що відгуки містять послідовність обговорення характеристик ноутбуків із мітками часу, наприклад обговорення щодо завантаження драйверів із відповідями служби підтримки виробника. Такий формат відгуків дає змогу проаналізувати послідовність взаємодії користувача із ІІС.

У процесі формування датасету з кожного відгуку із застосуванням ЛЛМ вилучено набір концептів, зокрема кількість слів у відгуку, терміни, застосовані у відгуку, числові значення показників у відгуку. Також відгук користувача структурується, тобто виокремлюються у відповіді переваги й недоліки і надається їх кількісна оцінка (кількість переваг і недоліків). На основі виділених концептів у відгуку розраховується його складність. Остання оцінюється як нормована сума кількості термінів, числових параметрів, а також переваг і недоліків. За складністю відгуку оцінюється рівень підготовки користувача. Застосовано емпіричні порогові правила. Отриманий датасет містить 44% фахових відгуків, 25% відгуків новачків та 31% відгуків досвідчених користувачів. Фахові відгуки містять не менше ніж п’ять технічних термінів і зазвичай наданих для ігрових ноутбуків. Покупцями цієї категорії комп’ютерів є, як правило, геймери з технічною підготовкою, які залишають розгорнуті відгуки. Ці відгуки, наприклад, містять виміри температур, fps, порівняння конфігурацій комп’ютера тощо. Новачки залишають короткі емоційні оцінки, наприклад "рекомендую", "купив і не шкодую" тощо. Такі оцінки не містять технічних термінів. Досвідчені користувачі залишають відгуки середньої довжини, що містять типові сценарії використання ноутбуків, наприклад "для роботи й навчання", "підходить для ігор". Однак і ці відгуки не мають технічних термінів.

Нейромережна частина ментальної моделі відтворює ключові ознаки рішення ("продуктивність", "автономність", "екран", "охолодження" тощо) в латентний простір ознак, як було продемонстровано в роботі [10].

Символьна частина формується через проміжний шар із нейромережної [10, 11] і може бути безпосередньо інтерпретована й проаналізована.

Оцінки складності отриманої ментальної моделі залежно від рівня підготовки користувачів подано в табл. 2.

Онтологія рішення ІІС (рекомендованого продукту в системі електронної комерції) за результатами експерименту містить 12 прийнятних для користувача концептів (властивостей рекомендованого продукту), зокрема: "продуктивність", "fps", "TDP", "автономність", "охолодження", "шум". Кількість концептів залежить від рівня підготовки (типу) користувача.

Таблиця 2. *Оцінки структурної складності нейросимвольної ментальної моделі*

Оцінка	***u_нов***	***u_пр***	***u_фах***
V^+ – прийняті концепти	4	7	12
V^- – негативні концепти	8	5	0
$\|G\|$ – кількість дуг каузального графа	3	6	18
$\Delta(M_i, u)$ – рівень деталізації моделі	0,30	0,60	1,00

Безрозмірний коефіцієнт $\Delta(M_i, u)$ має значення в діапазоні від 0 до 1 та кодує рівень деталізації відповідно до рівня компетентності користувача. Загальна ідея щодо вибору значень цього коефіцієнта полягає в такому. Користувач-новачок сприймає лише ключові поняття, які характеризують рекомендований товар або послугу в системі електронної комерції. Аналіз відгуків демонструє, що набір таких ключових понять становить приблизно 30% повного набору концептів. Просунутий користувач, крім базових понять, бере до уваги й сценарії застосування рекомендованого рішення, що зрештою становить понад половину набору концептів. А користувач-фахівець оперує концептами, що відтворюють усі, зокрема й технічні, властивості рекомендованого продукту.

З огляду на різний рівень деталізації каузальний граф у моделі нового користувача, яка має чотири концепти, містить лише три дуги, що відтворюють зв’язок між "продуктивність", "охолодження", "шум", "автономність". Граф у моделі фахового користувача містить, наприклад, дуги "DDR5 – {fps, теплопакет}", "CPU → {fps, температура, шум}".

Значення адаптивного порогу й темпоральної узгодженості, що встановлюють можливості персоналізації ментальної моделі, подано в табл. 3.

Адаптивний поріг $\theta(u)$ задає мінімальне значення подібності між концептом і послідовністю взаємодії користувача із ІІС.

Базове значення коефіцієнта 0,85 було використано в попередніх роботах для побудови нейросимвольної моделі. Таке значення дає змогу не перевантажувати новачка додатковими деталізованими концептами, що відтворюють

ISSN 3083-7715 (online) Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188)
ISSN 3083-7650 (0135-1710)(print) Management Information System and Devices. 2026. No. 1 (188)

властивості рішення ІІС. Мінімальне значення коефіцієнта становить 0,72. Експерименти на різнотипних наборах даних продемонстрували, що зниження коефіцієнта менше ніж 0,7–0,72 призводить до долучення несуттєвих концептів у ментальну моделі. Відповідно, за умови значення $\theta(u) = 0{,}81$ відбираються лише концепти, що відтворюють очевидні властивості рішення, а в разі $\theta(u) = 0{,}72$ в ментальну модель долучаються технічні властивості рішення.

Таблиця 3. *Персоналізація ментальної моделі*

Параметр	*u_нов*	*u_пр*	*u_фах*
$\theta(u)$	0,81	0,77	0,72
Зниження θ щодо базового	−0,04	−0,08	−0,13
TCI	1,0	0,97	0,95
Зміна TCI щодо базового значення	+0,33	+0,3	+0,28

Показник TCI (*Temporal Consistency Index*) визначає частку дуг каузального графа, що відповідає темпоральним обмеженням. Значення 1,0 для новачків демонструє, що модель містить лише ключові причинно-наслідкові зв'язки, в яких причина передує наслідку. Значення 0,95 для фахових користувачів свідчить, що одна з дуг може не задовольняти темпоральні обмеження. Значення приросту TCI щодо базового відтворює покращення моделі внаслідок переходу від фіксованого $\theta(u) = 0{,}85$ до адаптивного порогу. Порівняння отриманих коефіцієнтів (від 0,34 до 0,28) показує, що фіксований поріг не забезпечує деталізацію ментальної моделі насамперед для користувачів-новачків. Тобто такий поріг приводить до долучення до моделі технічних концептів, які не є релевантними для цих користувачів.

Крім цього, необхідно зауважити, що під час експерименту значення темпоральної узгодженості для методів LIME і SHAP визначались на основі порівняння з каузальним графом розробленої моделі.

Порівняння результатів роботи запропонованого фреймворку з позицій формування персоналізованих пояснень і наявних методів побудови пояснень за оцінкою TCI зображено на рис. 1.

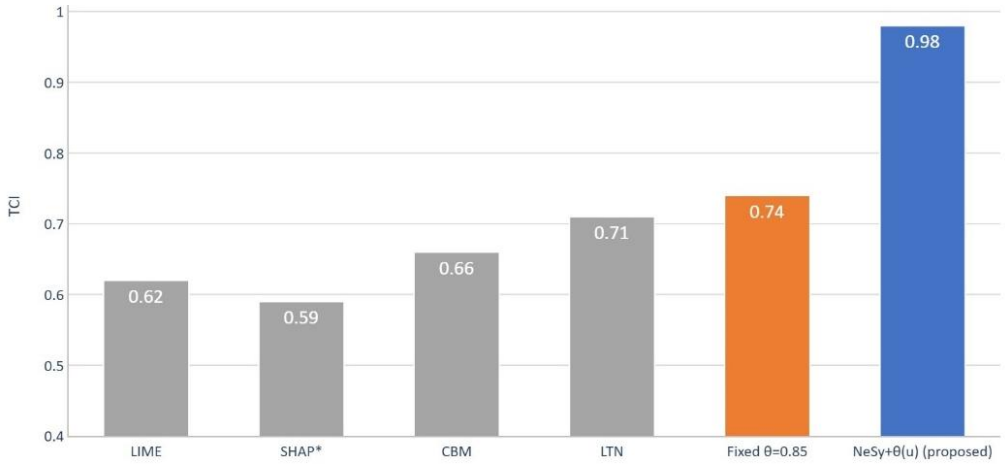

Рис. 1. Порівняльна оцінка TCI у процесі побудови ментальної моделі

Порівняння виконано в двох аспектах. По-перше, встановлено наявність суперечностей у поясненні з огляду на темпоральні залежності. По-друге, проаналізовано застосування розроблених принципів у процесі побудови пояснень.

Методи LIME і SHAP формують локальні пояснення без причинно-наслідкових залежностей. Вони беруть до уваги лише статистичні кореляції між властивостями рішення. Такий підхід приводить до зниження показника TCI до 0,62 для LIME та 0,59 для SHAP. Тобто користувач має сам відібрати 62% та 59% ознак, відповідно, із запропонованого цими методами пояснення.

Метод CBM (*Concept Bottleneck Models*) додає концепти, але не застосовує деталізації концептів, що й визначає показник TCI на рівні 0,66. Метод LTN (*Logic Tensor Networks*) використовує темпоральні оператори для формування обмежень, унаслідок чого й підвищується значення темпоральної узгодженості. Однак метод не персоналізує пояснення.

Обґрунтування запропонованих на рис. 1 результатів способом аналізу реалізованих цими

Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188) *ISSN 3083-7715 (online)*
Management Information System and Devices. 2026. No. 1 (188) *ISSN 3083-7650 (0135-1710)(print)*

методами принципів побудови ментальних моделей подано в табл. 4.

Як видно з табл. 4, перші п'ять методів не використовують повний набір принципів, що й зумовлює різницю в значеннях TCI.

Таблиця 4. *Використання принципів побудови ментальних моделей для персоналізації пояснень*

Метод	Реалізовані принципи
LIME	Відповідність структури (1), множинність (2)
SHAP	Відповідність структури (1), множинність (2)
CBM	Відповідність структури (1), множинність (2), неповнота (3)
LTN	Відповідність структури (1), множинність (2), неповнота (3), доповнення (4), темпоральна узгодженість (6)
NeSy з фіксованим порогом $\theta^* = 0,85$	Відповідність структури (1), множинність (2), неповнота (3), доповнення (4)
NeSy з адаптивним порогом $\theta(u)$	Відповідність структури (1), множинність (2), неповнота (3), доповнення (4), деталізація (5), темпоральна узгодженість (6), поведінкове виведення (7)

Обговорення результатів

Експериментально перевірені можливості розробленого нейросимвольного фреймворку зумовлені розбіжністю у поставці завдання, коли замість побудови типового пояснення для всіх користувачів формується персоналізована ментальна модель, поточна структура якої визначається за принципом деталізації згідно з рівнем підготовки користувача. Тобто кількість характеристик рішення і специфіка застосованих в описі рішення концептів залежать від рівня компетентності користувача інтелектуальної системи.

Застосування принципів деталізації, темпоральної узгодженості та поведінкового виведення підвищує релевантність отриманої моделі в темпоральному аспекті. Така перевага зумовлена тим, що використання заданого рівня деталізації обмежує множину можливих дуг у каузальному графі, а перевірка темпоральної узгодженості приводить до вилучення потенційних причинно-наслідкових залежностей, у яких наслідок передує причині.

Обмеження цього фреймворку пов'язані з особливостями подання вхідних даних. Зокрема українськомовні відгуки в системі електронної комерції містять переважно короткі емоційні оцінки, водночас можна використовувати різні мови в межах одного відгуку, що утруднює визначення рівня підготовки користувача.

Ключовою особливістю методу є повторюваність персоналізованого пояснення, тобто за незмінних вхідних даних у межах фреймворку буде створено одну й ту саму ментальну модель, що забезпечує відповідність вимогам щодо прозорості ІІС та можливість масштабування для платформ, які генерують значну кількість відгуків користувачів.

Напрями подальших досліджень пов'язані з розширенням онтології, що визначає характеристики рішення, а також з оцінюванням зрозумілості отриманих моделей на основі аналізу відгуків користувачів.

Висновки

Запропоновано нейросимвольний фреймворк формування ментальної моделі, який інтегрує концептуальне подання, систему принципів побудови, а також нейросимвольну реалізацію ментальної моделі. Це дає змогу сформувати персоналізовані пояснення щодо рішення інтелектуальної системи з огляду на поточний рівень підготовки користувача.

Удосконалено систему принципів побудови ментальних моделей рішення інтелектуальної системи, яка відрізняється від наявної принципами деталізації, темпоральної узгодженості та поведінкового виведення ментальної моделі. Це дає змогу в межах нейросимвольного підходу формувати ментальну модель із заданим рівнем деталізації з можливістю верифікації отриманих причинно-наслідкових залежностей у темпоральному аспекті.

Виконана експериментальна перевірка розробленого фреймворку підтвердила суттєве підвищення темпоральної узгодженості каузальних залежностей порівняно з наявними методами, а також використання повної системи принципів побудови ментальних моделей для персоналізації пояснень щодо рішень інтелектуальних систем.

Конфлікт інтересів

Автори декларують, що не мають конфлікту інтересів, зокрема фінансового, особистого,

ISSN 3083-7715 (online) *Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188)*
ISSN 3083-7650 (0135-1710)(print) *Management Information System and Devices. 2026. No. 1 (188)*

авторського чи будь-якого іншого характеру, який міг би вплинути на дослідження, а також на результати, опубліковані в цій статті.

Фінансування

Дослідження проведено без фінансової підтримки.

Доступність даних

Рукопис не має пов’язаних даних.

Використання засобів штучного інтелекту

Автори підтверджують, що не застосовували технології штучного інтелекту для написання статті.

References

1. Regulation (EU) 2024/1689 (2024), "Artificial Intelligence Act: laying down harmonised rules on artificial intelligence", *Official Journal of the European Union*, OJ L, 12.7.2024. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

2. Regulation (EU) 2016/679 (2016), "General Data Protection Regulation: on the protection of natural persons with regard to the processing of personal data and on the free movement of such data", *Official Journal of the European Union*, L 119, P. 1–88.

3. Johnson-Laird, P. N. (1983), Mental models: Towards a cognitive science of language, inference, and consciousness, Harvard University Press, Cambridge, 513 p.

4. Bansal, G., Nushi, B., Kamar, E., Weld, D. S., Lasecki, W. S., Horvitz, E. (2019), "Updates in human-AI teams: Understanding and addressing the performance/compatibility tradeoff", *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, No. 1, P. 2429–2437. DOI: https://doi.org/10.1609/aaai.v33i01.33012429

5. Brown, O., Power, N., Gore, J. (2024), "Cognitive task analysis: Eliciting expert cognition in context", *Organizational Research Methods*. DOI: https://doi.org/10.1177/10944281241271216

6. Sarker, M. K., Zhou, L., Eberhart, A., Hitzler, P. (2021), "Neuro-symbolic artificial intelligence: Current trends", *AI Communications*, Vol. 34, No. 3, P. 197–209. DOI: https://doi.org/10.3233/AIC-210084

7. Hitzler, P., Eberhart, A., Ebrahimi, M., Sarker, M. K., Zhou, L. (2022), "Neuro-symbolic approaches in artificial intelligence", *National Science Review*, Vol. 9, No. 6. DOI: https://doi.org/10.1093/nsr/nwac035

8. Chalyi, S. F., Leshchynska, I. O. (2023), "Conceptual mental model of explanation in an artificial intelligence system", *Bulletin of the National Technical University “KhPI”. Series: System Analysis, Control and Information Technology*, No. 1 (9), P. 70–75. DOI: https://doi.org/10.20998/2079-0023.2023.01.11

9. Chalyi, S. F., Leshchynska, I. O. (2024), "Principles of constructing mental models of decisions for an external user in the task of generating explanations in an intelligent system", *Automated Control Systems and Instruments of Automation*, No. 181, P. 82–90. DOI: https://doi.org/10.30837/0135-1710.2024.181.082

10. Chalyi, S. F., Leshchynska, I. O. (2026), "Integrated neuro-symbolic architecture of users’ mental models for personalised explanations of intelligent system decisions", *Aerospace Technic and Technology*, No. 1 (208), P. 133–140. DOI: https://doi.org/10.32620/aktt.2026.1.12

11. Chalyi, S. F., Leshchynska, I. O. (2025), "Method of constructing a neuro-symbolic representation of the mental model of an intelligent system’s decision", *Bionics of Intelligence*, No. 2 (103), P. 108–115. DOI: https://doi.org/10.30837/bi.2025.2(103).14

12. Longo, L., Brcic, M., Cabitza, F., Choi, J., Confalonieri, R., Del Ser, J. et al. (2024), "Explainable Artificial Intelligence (XAI) 2.0: A manifesto of open challenges and interdisciplinary research directions", *Information Fusion*, Vol. 106, P. 102301. DOI: https://doi.org/10.1016/j.inffus.2024.102301

13. 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*, San Francisco, USA, August 13–17, P. 1135–1144. DOI: https://doi.org/10.1145/2939672.2939778

14. Lundberg, S. M., Lee, S.-I. (2017), "A unified approach to interpreting model predictions", *Advances in Neural Information Processing Systems*, Vol. 30, P. 4765–4774.

15. Koh, P. W., Nguyen, T., Tang, Y. S., Pierson, E., Koh, J., Liang, P. (2020), "Concept bottleneck models", *Proceedings of the 37th International Conference on Machine Learning (ICML 2020)*, PMLR, Vol. 119, P. 5338–5348.

16. Badreddine, S., d’Avila Garcez, A., Serafini, L., Spranger, M. (2022), "Logic Tensor Networks", *Artificial Intelligence*, Vol. 303, P. 103649. DOI: https://doi.org/10.1016/j.artint.2021.103649

17. Manhaeve, R., Dumancic, S., Kimmig, A., Demeester, T., De Raedt, L. (2018), "DeepProbLog: Neural probabilistic logic programming", *Advances in Neural Information Processing Systems*, Vol. 31, P. 3753–3763.

18. Mao, J., Gan, C., Kohli, P., Tenenbaum, J. B., Wu, J. (2019), "The neuro-symbolic concept learner: Interpreting scenes, words, and sentences from natural supervision", *Proceedings of the 7th International Conference on Learning Representations (ICLR 2019)*, New Orleans, USA, May 6–9.

19. Miller, G. A. (1956), "The magical number seven, plus or minus two: Some limits on our capacity for processing information", *Psychological Review*, Vol. 63, No. 2, P. 81–97. DOI: https://doi.org/10.1037/h0043158

Автоматизовані системи управління та прилади автоматики. 2026. № 1 (188) *ISSN 3083-7715 (online)*
Management Information System and Devices. 2026. No. 1 (188) *ISSN 3083-7650 (0135-1710)(print)*

20. Cowan, N. (2001), "The magical number 4 in short-term memory: A reconsideration of mental storage capacity", *Behavioral and Brain Sciences*, Vol. 24, No. 1, P. 87–114. DOI: https://doi.org/10.1017/S0140525X01003922

Received (Надійшла) 11.02.2026
Accepted for publication (Прийнята до друку) 05.03.2026
Publication date (Дата публікації) 12.03.2026

Відомості про авторів / About the Authors

Чалий Сергій Федорович – доктор технічних наук, професор, Харківський національний університет радіоелектроніки, професор кафедри інтелектуальних управляючих систем; Харків, Україна;
Serhii Chalyi – Doctor of Technical Sciences, Professor, Kharkiv National University of Radio Electronics, Professor of the Intelligent Control Systems Department; Kharkiv, Ukraine;
e-mail: serhii.chalyi@nure.ua
ORCID ID: https://orcid.org/0000-0002-9982-9091

Лещинська Ірина Олександрівна – кандидат технічних наук, доцент, Харківський національний університет радіоелектроніки, доцент кафедри програмної інженерії; Харків, Україна;
Iryna Leshchynska – PhD (Engineering Sciences), Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor at the Software Engineering Department; Kharkiv, Ukraine;
e-mail: iryna.leshchynska@nure.ua
ORCID ID: https://orcid.org/0000-0002-8737-4595

A NEUROSYMBOLIC FRAMEWORK FOR FORMING MENTAL MODELS OF INTELLIGENT SYSTEM DECISIONS

The subject of the article *is the process of constructing mental models of intelligent system decisions for users with varying levels of competence and mapping their interaction sequences with intelligent information systems into structured mental representations.* ***The purpose of the work*** *is to develop a neurosymbolic approach integrating a conceptual mental model, a system of construction principles, and a neurosymbolic implementation of the mental model to enable personalized explanation generation in intelligent systems.* ***Tasks:*** *development of a neurosymbolic framework for constructing mental models of intelligent system decisions; extension of the system of principles to support explanation personalization based on user competence level; experimental verification of the neurosymbolic framework for constructing mental models of intelligent system decisions.* ***Research methods*** *are based on neurosymbolic artificial intelligence combining machine learning with causal symbolic reasoning, the application of linear temporal logic operators for verifying causal dependencies, and an adaptive concept selection mechanism based on user competence level.* ***Results achieved.*** *A neurosymbolic framework for constructing mental models of intelligent system decisions has been developed, which includes: a mental model integrating sets of positive and negative decision properties, a causal graph of cause-and-effect dependencies, and linguistic labels for natural-language interpretation of the model structure; an extended system of mental model construction principles that supplements the existing principles of correspondence, multiplicity, incompleteness, and complementation with the new principles of detailing, temporal consistency, and behavioral inference; an adaptive concept selection threshold function that determines the detailing level of the mental model according to the competence level of a specific user. It has been formally proved that the existing four-principle system is a necessary but not sufficient condition for the correctness of a neurosymbolic mental model. Experimental verification confirmed a substantial improvement in the temporal consistency of causal dependencies compared to existing methods.* ***Conclusions.*** *The proposed framework enables construction of personalized mental models with verified causal dependencies, satisfies the transparency requirements of artificial intelligence models, and provides scalable explanation personalization for intelligent systems.*

Keywords: *neurosymbolic artificial intelligence; mental model; intelligent information system; explainable AI; personalization; temporal consistency; causal reasoning; explanation generation.*

Бібліографічні описи / Bibliographic descriptions

Чалий С. Ф., Лещинська І. О. Нейросимвольний фреймворк формування ментальних моделей рішень інтелектуальних систем. *Автоматизовані системи управління та прилади автоматики*. 2026. № 1 (188). С. 98–107. DOI: https://doi.org/10.30837/0135-1710.2026.188.098

Chalyi, S., Leshchynska, I. (2026), "A neurosymbolic framework for forming mental models of intelligent system decisions", *Management Information System and Devices*, No. 1 (188), P. 98–107. DOI: https://doi.org/10.30837/0135-1710.2026.188.098