

**С. Ф. ЧАЛИЙ**, доктор технічних наук, професор, Харківський національний університет радіоелектроніки, професор кафедри інформаційних управляючих систем, м. Харків, Україна; e mail: serhii.chalyi@nure.ua, ORCID: <https://orcid.org/0000-0002-9982-9091>

**І. О. ЛЕЩИНЬСЬКА**, кандидат технічних наук (PhD), доцент, Харківський національний університет радіоелектроніки, доцент кафедри програмної інженерії; м. Харків, Україна; ; e-mail: iryna.leshchynska@nure.ua, ORCID: <https://orcid.org/0000-0002-8737-4595>

## ЕКСТЕРНАЛІЗАЦІЯ НЕЯВНИХ ЗНАТЬ В МЕНТАЛЬНІЙ МОДЕЛІ КОРИСТУВАЧА СИСТЕМИ ШТУЧНОГО ІНТЕЛЕКТУ

Предметом дослідження є процеси формування ментальної моделі користувача в системах штучного інтелекту. Побудова такої моделі пов'язана з вирішенням проблеми непрозорості та незрозумілості процесу прийняття рішень в таких системах для кінцевих користувачів. Щоб вирішити цю проблему, користувач системи має отримати пояснення щодо отриманого рішення. Пояснення повинно враховувати сприйняття рішення та процесу його прийняття користувачем, що формалізується в рамках ментальної моделі користувача. Ментальна модель враховує використання користувачем явних та неявних знань, останні з яких зазвичай не мають формального представлення. Екстерналізація таких знань забезпечує їх перетворення у формальну форму. Метою роботи є розробка підходу до екстерналізації неявних знань на основі виділення патернів та каузальних залежностей для процесу прийняття рішень в інтелектуальній системі при побудові ментальної моделі користувача. Для досягнення мети вирішуються такі завдання: розробка ментальної моделі користувача системи штучного інтелекту, яка враховує як явні, так і неявні знання, та розробка підходу до екстерналізації неявних знань користувача системи штучного інтелекту. Запропоновано ментальну модель користувача системи штучного інтелекту, яка враховує як явні, так і неявні знання користувача. Модель враховує зв'язки між явними та неявними знаннями користувача щодо системи штучного інтелекту, процесу прийняття рішень, способу використання рішень та загальної концепції інтелектуальної системи. Це створює умови для екстерналізації неявних знань користувача та подальшого використання цих знань при формуванні пояснень щодо процесу прийняття рішень у системі штучного інтелекту. Запропоновано підхід до екстерналізації знань зі статистичного та семантичного шарів ментальної моделі користувача. У практичному плані підхід дозволяє перевести в явну форму умови та обмеження щодо формування та використання рішень у системі штучного інтелекту.

**Ключові слова:** ментальна модель, пояснення, система штучного інтелекту, зрозумілий штучний інтелект, ментальна модель, каузальні залежності, причинно-наслідкові залежності.

**Вступ.** Ментальна модель з позицій когнітивної психології визначається як набір взаємопов'язаних елементів знань, які відображають внутрішнє представлення людини щодо конкретної фізичної системи, інформаційного чи фізичного процесу [1, 2]. Вказана модель дає можливість користувачеві явно представити поведінку такої системи та зробити прогнози щодо результатів її роботи.

Ментальна модель користувача для систем пояснювального штучного інтелекту (ХАІ) визначає представлення щодо сприйняття користувачем процесу прийняття рішення та результатів роботи системи штучного інтелекту. Така модель визначає порядок взаємодії з інтелектуальною системою, а також можливості ефективного використання отриманого від системи рішення. Ментальна модель є основою для створення пояснень, які допомагають користувачеві зрозуміти поведінку і результат роботи системи штучного інтелекту (ШІ). Відповідно, використання ментальної моделі дає можливість сформулювати умови для обґрунтованого прийняття рішень користувачем [3].

Ментальна модель користувача дає можливість підвищити ефективність використання рішень системи ШІ, на основі підвищення довіри користувача шляхом забезпечення прозорості роботи системи. По-перше, якщо користувач розуміє, як система штучного інтелекту приймає рішення, то він більше довіряє її висновкам. По-друге, така довіра базується на зрозумілому, прозорому для користувача представленні процесів в

системі ШІ, що знижує ризик неправильного трактування отриманих результатів. По-третє, внаслідок прозорого представлення процесу прийняття рішення полегшується взаємодія між користувачем і системою за рахунок спрощення інтерфейсу. В-четвертих, ментальна модель спрощує застосування отриманих рішень користувачем, надаючи необхідну інформацію для аналізу і валідації представленої в системі ШІ рішення.

Таким чином, ментальна модель користувача має бути врахована при побудові пояснень в системі штучного інтелекту для вирішення проблеми непрозорості процесу прийняття рішень у таких системах з позиції користувача.

Пояснення, яке відповідає ментальній моделі користувача, забезпечує довіру користувача до отриманого в інтелектуальній системі рішення. Це, у свою чергу, дає можливість ефективно використати отримане рішення на практиці [3].

Для підвищення довіри користувача до рішення інтелектуальної системи використовуються два основні підходи: використання моделей, що є прозорими або інтерпретованими у відповідності до ментальної моделі користувача; використання пояснень для алгоритмів прийняття рішень, що не можуть бути інтерпретовані безпосередньо.

Інтерпретація процесу прийняття рішення відбувається з використанням як явних, так і неявних знань користувача. Концепція неявних знань вперше була запропонована в роботі Michael Polanyi [4]. На відміну

© С. Ф. Чалий, І. О. Лещинська, 2024



Дослідницька стаття: Цю статтю опубліковано видавництвом *НТУ «ХПІ»* у збірнику «Вісник Національного технічного університету «ХПІ» Серія: Системний аналіз, управління та інформаційні технології». Ця стаття поширюється за міжнародною ліцензією [Creative Commons Attribution \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/). **Конфлікт інтересів:** Автор/и заявив/или про відсутність конфлікту.



від явних, неявні знання зазвичай відображають персональний досвід користувача. Подальший розвиток концепції неявних знань, запропонований в роботі японських дослідників I. Nonaka, H. Takeuchi [5] показав, що їх формалізація пов'язана із суттєвими труднощами. Відповідно, екстерналізація неявних знань в ментальній моделі користувача системи штучного інтелекту, тобто трансформація цих знань в явну форму є актуальною задачею.

Аналіз останніх досліджень і публікацій.

Дослідження щодо формування ментальних моделей для користувачів систем штучного інтелекту в останні роки проводились згідно програми пояснювального штучного інтелекту, яка передбачає формування пояснень щодо процесу прийняття рішень в інтелектуальній системі [6]. Необхідність врахувати сприйняття процесу роботи системи штучного інтелекту на пояснення була обґрунтована в [7]. Психологічні аспекти сприйняття пояснень, які свідчать про актуальність побудови ментальної моделі користувача системи штучного інтелекту, розглянуті в роботах [8, 9]. Можливості відображення неявних знань у формі темпоральних залежностей [10], а також інтеграції темпорального та каузального аспектів для пояснення процесу прийняття рішення в інтелектуальній системі були розглянуті в [11-13].

Аналіз представлених публікацій показує, що при побудові ментальних моделей користувача системи штучного інтелекту використовується переважно описовий підхід, в рамках якого не приділяється суттєва увага переведенню в явну форму та подальшому використанню неявних знань.

Однак для ефективного використання систем штучного інтелекту, зокрема на основі побудови пояснень щодо процесу їх роботи, необхідно екстерналізувати, тобто перевести в явну форму неявні знання з тим, щоб відобразити розуміння користувачем ключових принципів роботи системи штучного інтелекту та ключових етапів прийняття рішення в такій системі. Зазначене свідчить про важливість вирішення задачі екстерналізації таких знань.

**Мета та задачі дослідження.** Мета роботи полягає у розробці підходу до екстерналізації неявних знань на основі виділення каузальних залежностей для процесу прийняття та використання рішення в інтелектуальній системі в рамках ментальної моделі користувача.

Для досягнення мети вирішуються такі задачі:

- розробка ментальної моделі користувача системи штучного інтелекту, що враховує як явні, так і неявні знання;
- розробка підходу до екстерналізації неявних знань користувача системи штучного інтелекту.

Представлення явних та неявних знань у ментальній моделі користувача системи штучного інтелекту.

Ментальні моделі створюються людьми при вирішенні практичних задач. При побудові таких моделей виконується концептуалізація задач, що мають бути вирішені людиною, а також віртуалізація систем, які підтримують вирішення задач, та артефактів

навколишнього середовища, які використовуються в цих задачах.

У процесі побудови ментальної моделі людини використовуються як явні, так і неявні знання.

Згідно концепції неявних знань [4, 5], останні важко задокументувати та передати вербально. Значна частина наших ментальних моделей ґрунтується на досвіді та навичках, які не можуть бути повною мірою виражені словесно та в подальшому задокументовані для передачі іншим людям. Тому зазвичай такі знання залишаються в рамках особистого досвіду людини. Їх важко передати іншим людям без попередньої трансформації.

На відміну від неявних знань, явні знання є структурованими, систематизованими та документованими. Явні знання можуть бути передані між людьми за допомогою існуючих каналів комунікації. Прикладами явних знань є документи, інструкції, наукові публікації, стандарти тощо. Такі знання можуть бути використані для побудови схожих ментальних моделей у широкого кола людей.

Однак неявні знання є ключовими у процесі інновацій та навчання, тому використання цих знань та їх трансформація в явну форму відіграє суттєву роль у побудові та розвитку ментальних моделей.

Порівняльну характеристику явних та неявних знань з позиції їх представлення у ментальній моделі наведено в табл. 1.

Таблиця 1 – Представлення явних та неявних знань у ментальній моделі користувача системи штучного інтелекту

Знання	Особливості
1. Неявні	– типові шаблони процесу прийняття рішень у визначеній предметній області; – обмежені знання щодо умов та обмежень для використання рішень системи штучного інтелекту;
2. Явні	– причинно-наслідкові залежності між діями процесу прийняття рішень в інтелектуальній системі; – знання щодо загальних концепцій, на яких базується процес прийняття рішення в системі штучного інтелекту.

Можна виділити такі ключові знання людини, які впливають на характеристики ментальної моделі:

- загальний досвід користувача у визначеній предметній галузі;
- попередня інформація або досвід роботи із системами у відповідній предметній галузі;
- ключові алгоритми обробки інформації користувача;
- структура представлення даних про предметну галузь у конкретного користувача.

Ментальні моделі мають еволюційний характер, адаптуються до нової інформації, знань та досвіду користувача, забезпечуючи ефективну взаємодію з системами штучного інтелекту та використання результатів, отриманих у таких системах.

Однак ментальні моделі мають ряд недоліків, пов'язаних із використанням неявних знань.

По-перше, зазвичай ментальна модель є неповною внаслідок неповноти знань людини щодо предметної області. По-друге, ментальна модель може включати надлишкові правила, які не відображають особливості предметної області. Ці правила можуть базуватися на попередньому досвіді користувача. Вони забезпечують довіру до системи, з якою працює користувач, але не мають практичної значущості.

Структурно ментальна модель користувача має враховувати концепцію і принципи роботи системи, функції, які виконує система, особливості процесу прийняття рішення, а також контекст, у якому це рішення приймається.

Таким чином, при побудові ментальної моделі користувача в системі пояснювального штучного інтелекту, необхідно використовувати:

- концептуальні знання щодо процесу прийняття рішень у визначеній предметній області;
- знання щодо особливостей використання отриманого в системі результату;
- патерни, що відображають процес прийняття рішень у системі;
- каузальні залежності, що визначають зв'язок між етапами процесу прийняття рішення та отримання результату.

Концептуальні знання щодо процесу прийняття рішень відображають загальний досвід користувача у предметній області. Знання щодо особливостей використання рішення інтелектуальної системи відображають досвід роботи зі схожими системами в предметній області. Патерни процесу прийняття рішень доповнюють алгоритми обробки інформації користувачем. Вони відображають зв'язки між даними та результатом у відповідності до структури представлення даних про предметну область конкретного користувача.

Таким чином, ментальна модель користувача має бути побудована на концептуальному, семантичному, статистичному та каузальному рівнях з використанням відповідних підходів, як представлено в табл. 2.

Структуру ментальної моделі, що враховує явні та неявні аспекти знань користувача, представлено на рис. 1.

Формально модель містить шари з неявними знанням: статистичний  $B$  та семантичний  $S$ , а також шари з явними знанням: каузальний  $C$  та концептуальний  $P$ .

$$M = \langle B, S, C, P \rangle. \quad (1)$$

Елементи явних та неявних знань у моделі пов'язані наступним чином. Кожний елемент неявних знань у статистичному шарі має відповідний елемент явних знань у каузальному шарі. Тобто кожний патерн  $b_i$  із шару  $B$  складається із підмножини каузальних залежностей  $C_i$  із шару  $C$ .

$$B = \{b_i\}, C = \{c_j\}: \forall b_i \in B \exists C_i \subseteq C. \quad (2)$$

Іншими словами, кон'юнкція каузальних залежностей відображає в явному вигляді патерн  $b_i$ :

$$C_i = \{c_{j,1}, c_{j,2}, \dots, c_{j,l} : c_{j,1} \wedge \dots \wedge c_{j,l} \leftrightarrow b_i\}. \quad (3)$$

Аналогічно, кожен елемент семантичного шару  $s_k$  має концептуальне обмеження у відповідному шарі  $p_l$ :

$$S = \{s_k\}, P = \{p_l\}: (\forall k) \exists l : s_k | p_l. \quad (4)$$

Таблиця 2 – Рівні ментальної моделі користувача системи штучного інтелекту

Рівень моделі	Особливості
1. Каузальний	<ul style="list-style-type: none"> <li>– каузальні залежності між відомими станами системи або діями процесу прийняття рішення в інтелектуальній системі;</li> <li>– залежності між вхідними даними та результатом системи штучного інтелекту, що визначають вплив вхідних даних на отриманий результат;</li> <li>– на даному рівні забезпечується пояснення з урахуванням явних знань користувача щодо призначення вхідних і проміжних даних інтелектуальної системи;</li> </ul>
2. Статистичний (пов'язаний із неявними знаннями щодо процесу прийняття рішення)	<ul style="list-style-type: none"> <li>– типові патерни процесу прийняття рішення;</li> <li>– патерни відображають відомі дані про проміжні стани процесу формування результату в системі штучного інтелекту;</li> <li>– на даному рівні забезпечуються умови для пояснення рішення системи штучного інтелекту на основі досвіду (зазвичай неявних алгоритмів) обробки інформації користувачем.</li> </ul>
3. Семантичний (пов'язаний із рішенням системи)	<ul style="list-style-type: none"> <li>– контекст вирішення задачі, представлений вхідними даними;</li> <li>– умови використання отриманого рішення з урахуванням відмінностей вхідних даних;</li> <li>– на даному рівні забезпечуються умови для пояснення практичної важливості отриманого рішення з урахуванням, у тому числі неявних знань користувача щодо вирішення схожих до цільової задач у визначеній предметній області;</li> </ul>
4. Концептуальний (загальні закономірності функціонування системи, що є незалежними від контексту)	<ul style="list-style-type: none"> <li>– загальна схема процесу прийняття рішень у форматі «вхідні дані – ключові кроки – результат»;</li> <li>– загальні обмеження щодо процесу прийняття рішень, що є типовими для різних варіантів контексту вирішення цільової задачі;</li> <li>– ключові умови отримання результату, що не є залежними від поточного контексту;</li> <li>– на даному рівні забезпечуються умови для пояснення щодо принципової можливості отримати релевантне рішення при вирішенні цільової задачі у вибраній предметній області.</li> </ul>



Рис. 1. Структура ментальної моделі користувача у аспекті явних та неявних знань

Згідно з розглянутим представленням ментальної моделі користувача в аспекті явних та неявних знань, екстерналізація неявних знань відбувається при переході від статистичного до каузального шару ментальної моделі, а також при переході від семантичного до концептуального шару.

Тобто формально екстерналізація представлена функціями  $F_{BC} : B \rightarrow C$  та  $F_{SP} : S \rightarrow P$ . Перша з них визначає спосіб відображення патернів  $b_i$ , що відображають неявні знання про процес прийняття рішення у семантичному шарі, на підмножину явних каузальних залежностей  $C_i$ . Слід зазначити, що типові патерни відображають темпоральні залежності між діями процесу прийняття рішення або між вхідними даними та результатом. Тобто неявні знання мають темпоральні характеристики, що обумовлює можливість формування явних каузальних залежностей на відповідному шарі. Тому екстерналізація знань відбувається шляхом виявлення темпоральних залежностей, які мають причинно-наслідковий характер. Друга функція

визначає спосіб  $s_k$  використання отриманого в інтелектуальній системі результату при обмеженнях  $p_l$  у концептуальному шарі. Тобто екстерналізація знань у даному випадку здійснюється на основі встановлення обмежень застосування за принципом «що не заборонено, те може бути використано». Такий підхід враховує неповноту інформації щодо можливих варіантів використання рішення системи штучного інтелекту.

**Підхід до екстерналізації знань для статистичного та семантичного шарів ментальної моделі.** Підхід базується на таких запропонованих принципах: можливісний; обмежуючий. Перший принцип полягає у визначенні та оцінці можливості перетворення неявної каузальної залежності у складі патерну на явну причинно-наслідкову залежність. Використання можливісного принципу направлено на абстрагування від контексту застосування патернів, оскільки можливість розраховується у порівнянні з альтернативними варіантами каузальних залежностей на тих же наборах вхідних або проміжних даних. Зокрема, для рекомендаційної системи причинно-наслідковий зв'язок може бути визначений на основі значення можливості впливу певних характеристик товару на отриману рекомендацію (наприклад для рекомендованого кондиціонера – вплив потужності, торгової марки, управління по wi-fi тощо на вибір запропонованої моделі). Для системи розпізнавання відео – через значення можливості впливу певних елементів зображення на кінцевий результат розпізнавання.

Обмежуючий принцип полягає у екстерналізації неявних способів використання рішення системи штучного інтелекту через сукупність обмежень на можливості такого застосування. Ці обмеження мають відповідати концепції використання системи штучного інтелекту у вибраній предметній області. Зокрема, для мовних моделей, що генерують відповіді користувачам систем хостингу, такі обмеження можуть бути пов'язані із можливістю використання певних специфічних термінів, сленгових слів, а також повнотою інформації щодо технічних характеристик системи хостингу.

У відповідності до представлених принципів, розроблений підхід містить фази екстерналізації знань у статистичному та семантичному шарах моделі.

Фаза 1. Екстерналізація статистичних знань.

Вхідними даними є вхідні та проміжні дані системи штучного інтелекту.

Етап 1. Структуризація вхідних даних.

На даному етапі формується набір атрибутів вхідних даних. Для кожного атрибуту формується кінцева множина значень даних.

Етап 2. Розрахунок можливостей впливу кожного значення атрибутів на кінцевий результат (або наступну дію процесу прийняття рішення). Розрахунок виконується згідно теорії можливостей.

Етап 3. Відбір підмножини можливих значень причин.

На даному етапі відбирається підмножина значень атрибутів із значеннями можливості, що перевищують

порогове. Додатково може бути враховано показник необхідності.

Фаза 2. Екстерналізація семантичних знань.

Вхідними на даній фазі є інформація про призначення інтелектуальної системи. Така інформація може бути сформована на основі аналізу відповідної документації з використанням мовних моделей. Також в якості вхідних даних виступають загальні характеристики отриманого в інтелектуальній системі рішення. Наприклад, у рекомендаційній системі такі характеристики надаються в описі рекомендованого товару, а також у неявній формі, у відгуках користувачів на куплений товар.

Етап 1. Формування множини концептуальних обмежень з використанням мовної моделі, адаптованої під відповідну предметну область.

Етап 2. Формування переліку застосувань на основі аналізу описів використання рішення у різних контекстах.

Етап 3. Оцінка відповідності обмеженням на основі обчислення необхідності використання рішення.

Оцінка необхідності виконується шляхом порівняння рішення з альтернативами, що не задовольняють обмеженням.

Розглянемо приклад практичного застосування даного підходу для рекомендаційної системи.

Рекомендується певний товар, наприклад кондиціонер. Ключові характеристики товару: потужність охолодження; торгова марка; строк гарантії; преміальні властивості (можливість підігріву повітря, доступ через інтернет і т. п.). Проміжні етапи роботи рекомендаційної системи не є доступними для користувача. Відповідно, патерн вибору є невідомим. У такому випадку встановлюється причинно-наслідковий зв'язок між значенням властивостей товару та рекомендацією. Тобто необхідно знайти відповідь на запитання «З-за яких властивостей товару система рекомендувала його кінцевому користувачеві?». Для відповіді на це питання розраховуються значення можливостей для характеристик кондиціонера. Тобто окремо розраховуються відсотки кондиціонерів даної потужності, даної торгової марки, з преміальними функціями, які були продані в поточному сезоні. Можливості впливу цих значень визначаються згідно теорії можливостей. Відбираються властивості з максимальною можливістю. Наприклад, відповідь на наведене питання може мати вигляд: «кондиціонер було рекомендовано тому, що він має преміальну функцію підігріву повітря». Тобто отримано явні знання, які складають каузальний шар ментальної моделі.

На фазі 2 визначаються обмеження практичного застосування рекомендованого кондиціонера. Зокрема, аналіз відгуків з використанням мовної моделі може показати, що кондиціонер погано підігріває повітря при температурах нижче  $-15$  градусів. Згідно концептуального шару, кондиціонер, на відміну від теплового насосу, орієнтований в першу чергу на охолодження. Не передбачено ефективний нагрів при дуже низькій температурі (наприклад, нижче  $-7$  градусів) навколишнього середовища. Тобто встановлюється значення необхідності для даного товару з урахуванням

можливостей використання альтернативних моделей кондиціонерів для підігріву при низьких температурах зовнішнього середовища. Необхідність визначається згідно теорії можливостей. За результатом встановлюється обмеження на ефективний (з високим ККД) підігрів повітря.

**Висновки.** Запропоновано ментальну модель користувача системи штучного інтелекту, що враховує як явні, так і неявні знання користувача.

Модель складається з статистичного, семантичного, каузального та концептуального шарів. Концептуальний шар містить знання щодо схеми, принципів та концепції використання системи штучного інтелекту. Семантичний шар включає знання щодо умов та обмежень використання результатів, отриманих у системі штучного інтелекту. Статистичний шар відображає патерни формування рішень інтелектуальної системи. Каузальний шар містить ключові залежності між вхідними даними та рішеннями, а також між етапами процесу формування рішень у системі штучного інтелекту. Запропонована модель враховує зв'язки між явними та неявними знаннями користувача щодо системи штучного інтелекту, процесу прийняття рішень, способу використання рішень та загальної концепції інтелектуальної системи. Це дозволяє екстерналізувати неявні знання користувача та використати екстерналізовані знання при формуванні пояснень щодо процесу прийняття рішень.

Запропоновано підхід до екстерналізації знань у рамках ментальної моделі користувача системи штучного інтелекту. Підхід містить фази екстерналізації знань із статистичного шару моделі, що відображають типові патерни процесів прийняття рішень у системі, а також екстерналізації знань семантичного шару, що відображають способи використання цього рішення. У практичному плані підхід дає можливість перевести в явну форму умови та обмеження щодо формування та використання рішень у системі штучного інтелекту.

#### Список використаної літератури

1. Devedzic V., Radovic D., Jerinic L. A Framework for Building Intelligent Manufacturing Systems. *IEEE Transactions on Systems Man and Cybernetics Part C (Applications and Reviews)*, 1999. Vol. 29. P. 422–439. DOI: 10.1109/5326.777077.
2. Norman D. A. Some observations on mental models. *Human-Computer Interaction: A Multidisciplinary Approach*, San Francisco, CA: Morgan Kaufmann Publishers Inc., 1987. P. 241–244.
3. Чалий С. Ф., Лещинська І. О. Концептуальна ментальна модель пояснення в системі штучного інтелекту. *Вісник Національного технічного університету «ХПІ»*. Серія: Системний аналіз, управління та інформаційні технології. Харків: НТУ «ХПІ», 2023. № 1 (9). С. 70–75.
4. Polanyi M. *The Tacit Dimension*. New York, NY: Doubleday, 1966. 108 p.
5. Nonaka I., Takeuchi H. *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press, 1995. 304 p.
6. Gunning D., Vorm E., Wang J., Turek M. DARPA'S Explainable artificial intelligence (XAI) Program: a Retrospective. *Applied AI Letters* 2021. Vol. 2, no. 4. P. 1–11. <https://doi.org/10.1002/ail2.61>
7. Miller T. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*. 2019. Vol. 267. P. 1–38.
8. Keith J. H., Robert G. M. *The Oxford Handbook of Thinking and Reasoning* Get access Arrow. Oxford University Press, 2012. 864 p.
9. Carey S. *The origin of concepts*. New York, NY: Oxford University Press, 2009. 608 p.

10. Chala O. Development of information technology for the automated construction and expansion of the temporal knowledge base in the tasks of supporting management decisions. *Technology audit and production reserves*. 2019. № 1/2(45). P. 9–14. DOI: 10.15587/2312–8372.2019.160205.
11. Chalyi S., Leshchynskyi V. Temporal-oriented model of causal relationship for constructing explanations for decision-making process. *Advanced Information Systems*, 2022. № 6 (3). P. 60–65. <https://doi.org/10.20998/2522–9052.2022.3.09>
12. Chalyi S., Leshchynskyi V. Temporal representation of causality in the construction of explanations in intelligent systems. *Advanced Information Systems*. Kharkiv: NTU "KhPI", 2020. № 4 (3), P. 113–117.
13. Halpern J. Y., Pearl J. Causes and explanations: A structural–model approach. Part II: Explanations. *CoRR*. cs.AI/0208034. 10.1093/bjps/axi148.
4. Polanyi M. *The Tacit Dimension*. New York, NY: Doubleday, 1966. 108 p.
5. Nonaka I., Takeuchi H. *The Knowledge–Creating Company: How Japanese Companies Create the Dynamics of Innovation*. Oxford University Press, 1995. 304 p.
6. Gunning D., Vorm E., Wang J., Turek M. DARPA'S Explainable artificial intelligence (XAI) Program: a Retrospective, *Applied AI Letters*, 2021, vol. 2, no. 4, pp. 1–11. <https://doi.org/10.1002/ail2.61>
7. Miller T. Explanation in artificial intelligence: Insights from the social sciences, *Artificial. Intelligence*. 2019, vol. 267, pp. 1–38.
8. Keith J. H., Robert G. M. *The Oxford Handbook of Thinking and Reasoning Get access Arrow*. Oxford University Press, 2012. 864 p.
9. Carey S. *The origin of concepts*. New York, NY: Oxford University Press, 2009. 608 p.
10. Chala O. Development of information technology for the automated construction and expansion of the temporal knowledge base in the tasks of supporting management decisions. *Technology audit and production reserves*. 2019, no. 1/2(45), pp. 9–14. DOI: 10.15587/2312–8372.2019.160205.
11. Chalyi S., Leshchynskyi V. Temporal-oriented model of causal relationship for constructing explanations for decision-making process. *Advanced Information Systems*. 2022, no. 6(3), pp. 60–65. <https://doi.org/10.20998/2522–9052.2022.3.09>
12. Chalyi S., Leshchynskyi V. Temporal representation of causality in the construction of explanations in intelligent systems. *Advanced Information Systems*. Kharkiv, NTU "KhPI" Publ., 2020, no. 4 (3), pp. 113–117.
13. Halpern J. Y., Pearl J. Causes and explanations: A structural–model approach. Part II: Explanations. *CoRR*. cs.AI/0208034. 10.1093/bjps/axi148.

## References (transliterated)

1. Devedzic V., Radovic D., Jerinic L. A Framework for Building Intelligent Manufacturing Systems. *IEEE Transactions on Systems Man and Cybernetics Part C (Applications and Reviews)*, 1999, vol. 29, pp. 422–439. DOI: 10.1109/5326.777077.
2. Norman, D. A. Some observations on mental models. *Human-Computer Interaction: A Multidisciplinary Approach*, San Francisco, CA, Morgan Kaufmann Publishers Inc., 1987, pp. 241–244.
3. Chalyi S. F., Leshchynska I. O. Kontseptualna mentalna model poiasnennia v systemi shtuchnoho intelektu. *Visnyk Natsionalnoho tekhnichnoho universytetu «KhPI»*. Seriya: Systemnyi analiz, upravlinnia ta informatsiini tekhnologii [Bulletin of the National Technical University "KhPI": a collection of scientific papers. Series: System analysis, management and information technology]. Kharkiv, NTU "KhPI" Publ., 2023, no. 1 (9), pp. 70–75.
11. Chalyi S., Leshchynskyi V. Temporal-oriented model of causal relationship for constructing explanations for decision-making process. *Advanced Information Systems*. 2022, no. 6(3), pp. 60–65. <https://doi.org/10.20998/2522–9052.2022.3.09>
12. Chalyi S., Leshchynskyi V. Temporal representation of causality in the construction of explanations in intelligent systems. *Advanced Information Systems*. Kharkiv, NTU "KhPI" Publ., 2020, no. 4 (3), pp. 113–117.
13. Halpern J. Y., Pearl J. Causes and explanations: A structural–model approach. Part II: Explanations. *CoRR*. cs.AI/0208034. 10.1093/bjps/axi148.

Надійшла (received) 10.05.2024

UDC 004.891.3

**S. F. CHALYI**, Doctor of Technical Sciences, Full Professor, Kharkiv National University of Radio Electronics, Professor of the Department of Information Control System, Kharkiv; e mail: serhii.chalyi@nure.ua, ORCID: <https://orcid.org/0000-0002-9982-9091>

**I. O. LESHCHYNSKA**, Candidate of Technical Sciences (PhD), Associate Professor, Kharkiv National University of Radio Electronics, Associate Professor at the Department of Software Engineering доцент кафедри програмної інженерії, Kharkiv; e-mail: iryna.leshchynska@nure.ua, ORCID: <https://orcid.org/0000-0002-8737-4595>

## EXTERNALIZATION OF TACIT KNOWLEDGE IN THE MENTAL MODEL OF A USER OF AN ARTIFICIAL INTELLIGENCE SYSTEM

The subject of the study is the processes of forming the user's mental model in artificial intelligence systems. The construction of such a model is associated with solving the problem of opacity and incomprehensibility of the decision-making process in such systems for end users. To solve this problem, the system user needs to receive an explanation of the obtained decision. The explanation should take into account the user's perception of the decision and the decision-making process, which is formalized within the user's mental model. The mental model considers the user's use of explicit and implicit knowledge, the latter of which usually lacks formal representation. The externalization of such knowledge ensures its transformation into a formal form. The aim of the work is to develop an approach to the externalization of implicit knowledge based on identifying patterns and causal dependencies for the decision-making process in an intelligent system when constructing the user's mental model. To achieve this goal, the following tasks are solved: developing a user's mental model of an artificial intelligence system that takes into account both explicit and implicit knowledge and developing an approach to the externalization of implicit knowledge of the user of the artificial intelligence system. A user's mental model of an artificial intelligence system that accounts for both explicit and implicit knowledge of the user is proposed. The model considers the connections between the user's explicit and implicit knowledge regarding the artificial intelligence system, the decision-making process, the method of using the decisions, and the general concept of the intelligent system. This creates conditions for the externalization of the user's implicit knowledge and the subsequent use of this knowledge in forming explanations regarding the decision-making process in the artificial intelligence system. An approach to the externalization of knowledge from the statistical and semantic layers of the user's mental model is proposed. In practical terms, the approach makes it possible to translate into explicit form the conditions and constraints regarding the formation and use of decisions in the artificial intelligence system.

**Keywords:** mental model, explanation, artificial intelligence system, explainable artificial intelligence, mental model, causal dependencies, cause-and-effect relationships.

Повні імена авторів / Author's full names

**Автор 1 / Author 1:** Чалий Сергій Федорович, Chalyi Serhii Fedorovych

**Автор 2 / Author 2:** Лещинська Ірина Олександрівна, Leshchynska Irina Oleksandrivna