УДК 004.8:001.891
DOI https://doi.org/10.15589/znp2025.3(501).21

TAXONOMY OF ARTIFICIAL INTELLIGENCE AUTONOMY IN SCIENTIFIC RESEARCH

ТАКСОНОМІЯ АВТОНОМНОСТІ ШТУЧНОГО ІНТЕЛЕКТУ В НАУКОВИХ ДОСЛІДЖЕННЯХ

Yuliia G. Loboda
loboda@onua.edu.ua ORCID: 0000-0001-7083-552X
Olena G. Trofymenko
trofymenko@onua.edu.ua
ORCID: 0000-0001-7626-0886
Nataliia I. Loginova
ORCID: 0000-0002-9475-6188
loginova@onua.edu.ua
Olexander V. Zadereyko
zadereyko@onua.edu.ua
ORCID: 0000-0003-0497-9861
Anatoliy Yu. Gaida[2]
cetus@ukr.net
ORCID: 0000-0002-8217-5044

Ю. Г. Лобода,
канд. пед. наук, доцент
О. Г. Трофименко,
канд. техн. наук, доцент
Н. І. Логінова,
канд. пед. наук, доцент
О. В. Задерейко, канд. техн. наук, доцент
А. Ю. Гайда[2]**,**
канд. техн. наук

[1] National University «Odesa Law Academy», Odessa
[2] Admiral Makarov National University of Shipbuilding, Mykolaiv
[1] Національний університет «Одеська юридична академія», м. Одеса
[2] Національний університет кораблебудування імені адмірала Макарова, м. Миколаїв

Abstract. The article is devoted to the development of a systematized taxonomy of artificial intelligence (AI) autonomy levels in the context of scientific research, as well as to the definition of corresponding trust profiles for the safe integration of AI systems into research practice. The relevance of the topic is driven by the rapid growth of AI's role in knowledge generation, research automation, hypothesis formation, and result interpretation, which requires new approaches to autonomy classification and trust systems. The *purpose* of this article is to develop a five-level taxonomy of AI autonomy, tailored to the specific requirements of knowledge generation, critical thinking, and ethical responsibility in scientific activities. The research *methodology* involves the use of system analysis of up-to-date sources, comparative analysis of autonomy classifications across different domains, taxonomic modeling, and synthesis of interdisciplinary approaches from AI theory, trust psychology, and human-machine interaction. Both inductive and deductive methods are applied to generalize the characteristics of autonomy levels and to form trust profiles. The *results* include the development of a five-level taxonomy of AI autonomy, ranging from passive tools to fully autonomous scientific systems. For each level, a specific trust profile is defined, incorporating cognitive, affective, and behavioral components. Validation methods for trust are systematized according to the level of system autonomy: quantitative, qualitative, and mixed approaches. Particular attention is paid to the concept of trust calibration as a means of preventing over-trust or under-trust. *Scientific novelty*: for the first time, a taxonomy of AI autonomy tailored to the needs of scientific research is proposed, taking into account task complexity, decision transparency, and ethical aspects of responsibility. A conceptual model of multi-component trust toward autonomous scientific systems has been developed. *Practical significance*: the research findings establish a methodological framework for the ethical integration of AI into scientific practice, the establishment of validation standards for autonomous systems, and the preparation of researchers for safe and effective interactions with AI of varying autonomy levels.
Key words: artificial intelligence; scientific research; classification; autonomy; validation; agent systems; human-machine interaction.

Анотація. Статтю присвячено розробці систематизованої таксономії рівнів автономності штучного інтелекту в контексті наукових досліджень та визначенню відповідних профілів довіри для безпечної інтеграції систем

штучного інтелекту у дослідницьку практику. Актуальність теми зумовлена стрімким зростанням ролі штучного інтелекту у процесах генерації знань, автоматизації досліджень, формування гіпотез та інтерпретації результатів, що потребує розробки нових підходів до класифікації автономності та систем довіри. *Метою статті* є створення п'ятирівневої таксономії автономності штучного інтелекту, адаптованої до специфіки генерації знань, критичного мислення та етичної відповідальності в науковій діяльності. *Методика* дослідження полягає у використанні системного аналіз сучасних джерел, порівняльного аналізу класифікацій автономності в різних галузях, таксономічного моделювання, синтезу міждисциплінарних підходів з теорії штучного інтелекту, психології довіри та людино-машинної взаємодії. Застосовано індуктивні та дедуктивні методи для узагальнення характеристик рівнів автономності та формування профілів довіри. *Результатами* є створення п'ятирівневої таксономії автономності штучного інтелекту, від пасивного інструмента до автономної наукової системи. Для кожного рівня визначено специфічний профіль довіри, що включає когнітивні, афективні та поведінкові компоненти. Систематизовано методики валідації довіри відповідно до рівня автономності системи: кількісні, якісні та змішані. Окрему увагу приділено концепції калібрування довіри як засобу запобігання її завищенню або заниженню. *Наукова новизна* – вперше запропоновано адаптовану до потреб наукових досліджень таксономію автономності штучного інтелекту, яка враховує когнітивну складність завдань, прозорість рішень та етичні аспекти відповідальності. Розроблено концептуальну модель багатокомпонентної довіри до автономних наукових систем. *Практична значимість* полягає у тому, що результати дослідження створюють методологічну основу для етичної інтеграції штучного інтелекту в наукову практику, формування стандартів валідації автономних систем, а також для підготовки дослідників до безпечної й ефективної взаємодії з штучним інтелектом різних рівнів автономності.

Ключові слова: штучний інтелект; наукові дослідження; класифікація; автономність; валідація; агентні системи; людино-машинна взаємодія.

ПОСТАНОВКА ЗАДАЧІ

Упродовж останнього десятиліття штучний інтелект (ШІ) радикально трансформував ландшафт наукових досліджень. Сучасні системи ШІ поступово еволюціонували від базових інструментів статистичного аналізу для обробки даних до складних інтелектуальних агентів, здатних генерувати гіпотези, планувати експерименти, аналізувати результати та формулювати наукові висновки. Подібна трансформація зумовлює фундаментальну зміну ролі дослідника від повного контролю над науковим процесом до співпраці з автономними системами ШІ, які частково або повністю беруть на себе ключові функції наукової діяльності. Швидке впровадження ШІ на всіх етапах дослідження від пошуку літератури до моделювання та інтерпретації супроводжується низкою нових викликів. Зокрема постають питання: як класифікувати типи автономності ШІ в контексті дослідницької діяльності? які завдання можна безпечно делегувати системам з певним рівнем самостійності? якими мають бути критерії довіри до таких систем, особливо у випадках, коли їхні дії впливають на наукові висновки? Попри наявні класифікації автономності в окремих галузях наразі бракує моделей, адаптованих до потреб наукових досліджень. Наукова діяльність охоплює специфічні аспекти генерації знань, критичного мислення, етичної відповідальності, верифікації результатів. На відміну від автоматизації рутинних завдань, яка є типовим прикладом використанням ШІ для інших галузей, наукові дослідження вимагають глибшого підходу до класифікації автономності ШІ, з урахуванням прозорості міркувань, аргументації та контролю над ухваленням рішень. У цьому контексті актуальним є створення таксономії рівнів автономності ШІ у наукових дослідженнях та відповідних профілів довіри, які враховують когнітивну складність завдань, ступінь відповідальності систем і потребу в контролі.

АНАЛІЗ ОСТАННІХ ДОСЛІДЖЕНЬ І ПУБЛІКАЦІЙ

Проблема класифікації рівнів автономності та довіри до ШІ в наукових дослідженнях є предметом активного міждисциплінарного дискурсу. Аналіз сучасної літератури дозволяє виокремити два основні напрями досліджень: розробку таксономії автономності та вивчення механізмів формування довіри до систем ШІ. Спроби класифікувати ШІ за рівнем автономності здійснюються в різних галузях. Згідно з класифікацією, запропонованою в межах проєкту Європейського аналітичного центру Interface, виділено п'ять рівнів автономності агентів на основі ступеня людського контролю та ініціативності системи [1]. Подібний підхід застосовується у сфері медицини/ Так, Американська медична асоціація (AMA) розробила трирівневу таксономію ШІ для клінічного використання: assistive (допоміжний), augmentative (доповнюючий) та autonomous (автономний), де кожен рівень має різний ступінь втручання людини [2]. У сфері транспорту, зокрема автономного водіння, найбільш поширеною є класифікація за стандартом SAE International, яка має шість рівнів автономності: від нульового (повна залежність від водія) до п'ятого (повна автономність без потреби втручання людини) [3]. Такий підхід набув широкого визнання і використовується як основа для регуляторних політик та

тестування систем. Отже, у різних галузях спостерігається схожість підходів до класифікації автономності ШІ, що свідчить про загальну потребу в уніфікованих критеріях оцінки рівня самостійності ШІ та довіри до нього. Питання таксономії набули особливої актуальності з появою так званих агентних систем (Agentic AI). Сучасні дослідження визначають їх як системи, здатні незалежно виконувати повний цикл наукової діяльності: від генерації гіпотез і пошуку літератури до проєктування експериментів, аналізу результатів і формулювання висновків [4]. Як зазначають автори [5], такі системи не лише автоматизують окремі етапи дослідження, а й здатні ініціювати дії, які традиційно вважалися виключно сферою людської компетенції. Дані ідеї перекликаються з дослідженнями [6, 7], в яких розглядаються моделі довіри до систем із різним рівнем автономії. В них показано, що користувачі менш охоче довіряють напівавтономним і повністю автономним системам, оцінюючи їх не лише за результатами, а й за поведінковими та емоційними характеристиками. Окрему увагу проблематиці автономності та довіри у контексті системного аналізу приділено у дослідженнях [8, 9], де розглядаються інструменти ШІ та особливості їхнього впровадження в наукову діяльність.

Попри це, у сфері академічних досліджень досі відсутня єдина, адаптована до наукового контексту таксономія автономності. Часто застосовуються підходи з інших сфер, як-от: транспорт, оборона або медицина [10–12], що створює методологічну прогалину й ускладнює системну інтеграцію ШІ у дослідницьку діяльність.

Автономність ШІ тісно пов'язана з довірою користувача. Чим вищий рівень самостійності системи, тим більше значення мають пояснюваність, прозорість, відповідність очікуванням дослідника. Як зазначено в роботі [6], довіра до автономних систем залежить не лише від їх технічної ефективності, а й від зрозумілості дій ШІ для людини. Дослідження підкреслюють, що хоча пояснення рішень ШІ мають доведені переваги, їхній вплив у командному середовищі «людина – ШІ» ще недостатньо вивчений [13]. У цьому контексті важливим є попередження Королівського товариства про те, що надмірна автономність без відповідної прозорості може підірвати довіру до наукових результатів [14]. Довіра до ШІ залежить також і від антропоморфних ознак системи – мови, інтонацій, стилю взаємодії. Користувачі часто оцінюють такі системи не лише раціонально, а й емоційно [7]. Це підтверджують й автори дослідження [15], в якому встановлено вплив рівня автономності на когнітивну оцінку системи користувачем. Наразі відсутні узгоджені підходи до формування довіри до агентів ШІ, які діють як активні учасники дослідницького процесу. В роботі [16] наголошено: «довіритель має довіряти лише настільки, наскільки заслуговує довіри той, кому довіряють». Дане твердження лягло в основу концепції trusted autonomy, яка охоплює не лише технічну надійність, а й етичну, соціальну та психологічну відповідальність автономних систем [17]. Загалом сучасні публікації [18–20] підтримують ідею багатофакторної моделі довіри, яка охоплює особистісні характеристики користувача, його попередній досвід, контекст використання та рівень автономності.

Водночас зі зростанням автономності ШІ зростають і методологічні виклики. У дослідженні [21] вказано, що автономні агенти можуть змінити когнітивні та етичні межі науки, оскільки здатні ініціювати гіпотези без участі людської інтуїції, що створює ризики формального, а не змістовного обґрунтування результатів. Крім того, дослідники застерігають про те, що автономні системи можуть витіснити людське агентство, що потенційно спричиняє зниження мотивації до критичного мислення, відповідальності та етичної рефлексії [22].

ВІДОКРЕМЛЕННЯ НЕВИРІШЕНИХ РАНІШЕ ЧАСТИН ЗАГАЛЬНОЇ ПРОБЛЕМИ

Аналіз сучасних досліджень показав, що, попри наявність різних класифікацій автономності ШІ, більшість із них не адаптовані до специфіки академічного середовища. Довіра до автономних систем визначається рівнем пояснюваності, контекстом використання, індивідуальними особливостями користувача та характером взаємодії з системою. Відсутність прозорості та контрольованості в ухваленні рішень може підірвати достовірність наукових висновків. У цьому контексті зростання рівня автономності ШІ вимагає розробки нових методологічних підходів та етичних рамок для їх інтеграції у наукову практику.

МЕТА ДОСЛІДЖЕННЯ

Метою дослідження є створення п'ятирівневої таксономії автономності штучного інтелекту, адаптованої до специфіки наукових досліджень, та розробка відповідних профілів довіри для безпечної інтеграції систем ШІ у дослідницьку практику.

МЕТОДИ, ОБ'ЄКТ ТА ПРЕДМЕТ ДОСЛІДЖЕННЯ

Методи. Дослідження проводиться з використанням загальнонаукових методів дослідження, як-от: пошук, системний аналіз, спостереження, синтез, абстрагування та порівняння, абдукція, ідеалізація, пояснення та аналогія, історичний і логічний методи.

Об'єктом дослідження є системи штучного інтелекту різних рівнів автономності в контексті наукової діяльності.

Предметом дослідження є закономірності формування довіри до автономних систем ШІ залежно від рівня їх самостійності у виконанні дослідницьких завдань.

ОСНОВНИЙ МАТЕРІАЛ

Рівні автономності ШІ в наукових дослідженнях. Автономність ШІ в контексті наукових досліджень розглядають як здатність системи самостійно ухвалювати рішення, формулювати цілі та виконувати дії без безпосереднього втручання людини [1, 4]. На відміну від багатьох інших галузей, де автономність зводиться до автоматизації рутинних процесів, у наукових дослідженнях вона охоплює складні когнітивні завдання, такі як генерацію гіпотез, критичне осмислення результатів, верифікацію знань [5, 6]. Аналіз існуючих класифікацій автономності, зокрема, SAE для транспортної галузі [3], AMA – для медицини [1-2] демонструє фокус на ступені людського контролю. Проте наукова діяльність потребує розширених критеріїв, які враховують здатність до аргументації, прозорість процесу прийняття рішень, оцінки достовірності та відтворюваності результатів [7, 10]. Відсутність адаптованої моделі автономності для наукового контексту створює методологічну прогалину, що ускладнює впровадження ШІ у дослідницьку практику. У зв'язку з цим виникає потреба у створенні таксономії, яка відображатиме специфіку використання ШІ у наукових дослідженнях, від допоміжних інструментів до повноцінних автономних агентів, здатних діяти в ролі науковця.

На основі аналізу поточних досліджень і практик використання ШІ в науковому середовищі запропоновано п'ятирівневу таксономію автономності систем ШІ, яка враховує рівень когнітивної складності, характер контролю та розподіл відповідальності (табл. 1).

Запропонована модель ґрунтується на принципах когнітивної прогресії та поступової передачі відповідальності, що узгоджується з концепцією «довіреної автономії» (trusted autonomy) [16, 17]. Вона фіксує еволюцію від допоміжного інструмента до повністю автономного агента, здатного здійснювати дослідження без участі людини.

На рис. 1 подано ієрархічну структуру з п'яти рівнів автономності ШІ у наукових дослідженнях.

Наведена ієрархічна структура має поступову еволюцію від рівня 0 (Пасивний інструмент) до рівня 4 (Автономна наукова система). Кожен рівень характеризується підвищенням когнітивної складності та зменшенням людського контролю.

Перехід між рівнями автономності не є лінійним. Він залежить від низки факторів:

– когнітивної складності завдань, наприклад: від обробки табличних даних до критичної оцінки гіпотез;

– ступеня самостійності в ухваленні рішень: від прямого виконання команд до стратегічного планування;

– здатності до адаптації за умов невизначеності;

– розподілу відповідальності за дії та висновки;

– потреби в зовнішньому контролі [12, 14, 15].

Усі ці фактори мають не лише технічне, а й психологічне значення, адже вони впливають на рівень довіри, яку дослідник готовий делегувати системі ШІ.

Профілі довіри до ШІ в наукових дослідженнях. Формування довіри до систем ШІ в дослідницькому контексті є складним, багатофакторним процесом. На відміну від прикладних сфер, де достатньо технічної надійності або ефективності, у наукових дослідженнях вирішальне значення мають прозорість дій, відтворюваність результатів та відповідність науковій логіці. Довіра до ШІ формується на перетині когнітивних, емоційних, соціальних та етичних чинників [18, 19].

На основі аналізу досліджень у галузях психології довіри, когнітивної ергономіки та взаємодії людини з автономними системами [7, 16, 23–26] виокремлюють три ключові компоненти довіри:

– когнітивна довіра – раціональна оцінка технічної компетентності системи, яка ґрунтується на показниках точності, стабільності та відтворюваності результатів [23];

Таблиця 1. Рівні автономності ШІ у наукових дослідженнях

Рівень	Назва	Основні характеристики	Приклади систем	Рівень відповідальності
0	Пасивний інструмент (Passive Tool)	Виконує явно задані команди, не проявляє ініціативи	Excel, SPSS, R (базовий), Zotero	Повна відповідальність дослідника
1	Активний аналітик (Active Analyst)	Виявляє закономірності, оптимізує параметри, надає рекомендації	AutoML, ImageJ з AI-модулями, ATLAS.ti	Дослідник формулює завдання та інтерпретує результати
2	Генератор гіпотез (Hypothesis Generator)	Генерує нові ідеї на основі знань і контексту	GPT-4, Claude, IBM Watson	Дослідник оцінює валідність запропонованих ідей
3	Автономний експериментатор (Autonomous Experimenter)	Планує, адаптує і виконує експерименти з мінімальним наглядом	Emerald Cloud Lab, AI for Chemistry	Система виконує експерименти, дослідник здійснює валідацію
4	Автономна наукова система (Autonomous Scientific System)	Виконує повний дослідницький цикл, включно з саморефлексією	Прототипи агентів типу «AI Scientist»	Основна відповідальність на системі, людина валідує

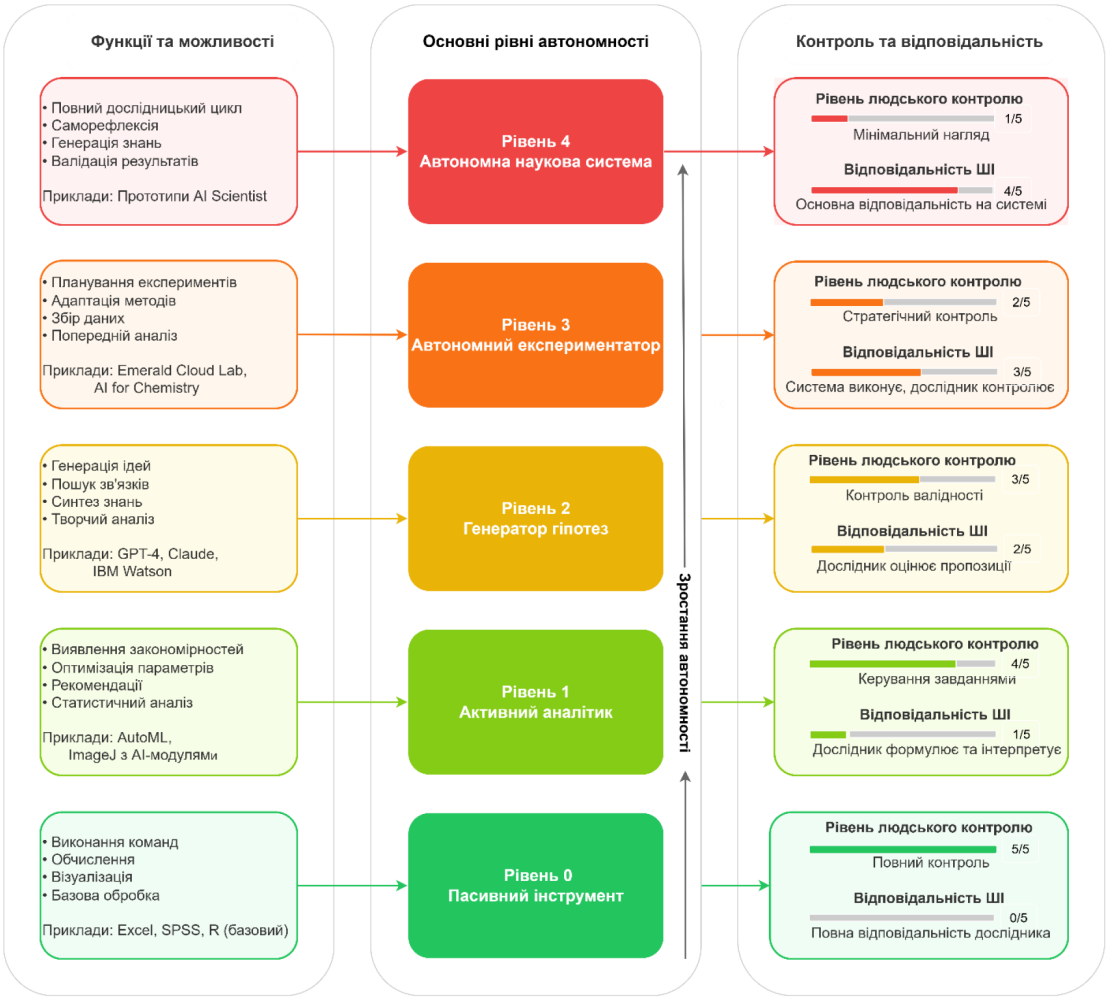

Рис. 1. Ієрархічна структура рівнів автономності ШІ у наукових дослідженнях

– афективна довіра – емоційне ставлення до системи, що формується через інтерфейс, до стиля комунікації ШІ, а також схильність користувача до прийняття нових технологій [7, 15];

– поведінкова довіра – готовність діяти відповідно до рішень або рекомендацій системи у реальних дослідницьких ситуаціях [24].

Кожному рівню автономності ШІ, від пасивного інструмента до автономної наукової системи, відповідає специфічний профіль довіри, який визначає характер взаємодії між дослідником і системою, розподіл відповідальності й контролю, вимоги до пояснюваності рішень та емоційне сприйняття системи (помічник, партнер, конкурент тощо).

На рис. 2 показано радарну діаграму профілів довіри для п'яти рівнів автономності ШІ. Кожен рівень характеризується унікальним поєднанням когнітивної, афективної та поведінкової довіри, що формує специфічний «відбиток довіри». Візуалізація демонструє, як зміна рівня автономності впливає на баланс різних компонентів довіри, та дозволяє дослідникам визначити оптимальний рівень автономності відповідно до їхніх потреб та готовності до співпраці з системою.

На рівні 0–1 довіра має інструментальний характер, дослідник не очікує ініціативи, система цілком підконтрольна. На рівні 2 виникає оціночна довіра – система генерує ідеї, які дослідник критично оцінює, зберігаючи за собою остаточне рішення. На рівнях 3–4 формується структурована довіра, коли дослідник сприймає ШІ як автономного агента, делегуючи йому певні функції, але зберігаючи стратегічний нагляд або функцію валідації.

Ефективна інтеграція ШІ у наукову діяльність вимагає не лише технічної перевірки його функцій, а й ретельної оцінки довіри до систем з боку дослідників. Згідно з дослідженнями [15, 16, 26] методи валідації довіри до ШІ можна класифікувати за трьома напрямами: кількісні, якісні та змішані.

1. *Кількісні підходи* базуються на об'єктивних метриках, які дозволяють вимірювати продуктивність і надійність систем:

– metrics-based validation – оцінювання точності, повноти, F1-score, стабільності результатів [27];

Таблиця 2. Профілі довіри до систем ШІ залежно від рівня автономності

Рівень автономності	Когнітивна довіра	Афективна довіра	Поведінкова довіра	Основні критерії валідації
0 Пасивний інструмент	Висока (повний контроль)	Нейтральна	Висока в межах функцій	Точність, стабільність, відсутність збоїв
1 Активний аналітик	Середня	Позитивна за UX	Обмежена, із верифікацією	Пояснюваність, верифікація, документація рішень
2 Генератор гіпотез	Низька-середня (критична оцінка)	Змінна	Обережна	Новизна, логічність, емпірична перевірка
3 Автономний експериментатор	Висока в технічному аспекті	Позитивна з досвідом	Висока (контроль висновків)	Відтворюваність, адаптивність, повна документація
4 Автономна наукова система	Потребує зовнішньої експертизи	Скептична	Мінімальна	Реплікація повного циклу, етична та наукова експертиза

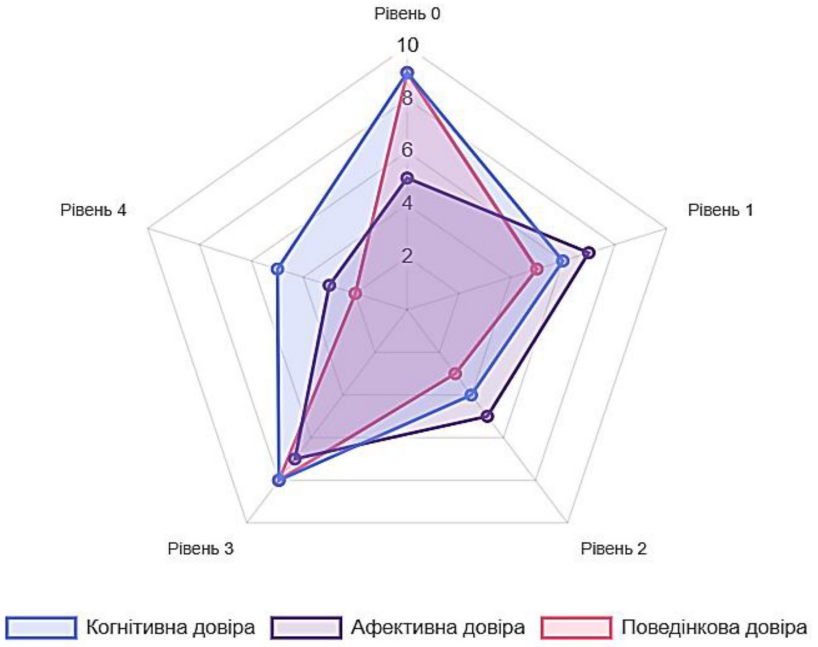

Рис. 2. Діаграма профілів довіри для п'яти рівнів автономності ШІ

– статистичне моделювання довіри – використання байєсівських підходів оцінки ймовірності правильності рішень [28];

– порівняльний аналіз – тестування систем на контрольних наборах із відомими результатами.

Такі методи особливо ефективні для рівнів 0–2, де головне – оцінити якість рішень і обґрунтованість висновків.

2. *Якісні підходи* орієнтовані на досвід користувача та соціальні аспекти:

– експертне оцінювання – перевірка гіпотез, згенерованих ШІ;

– феноменологічний аналіз – вивчення суб'єктивного досвіду взаємодії з системою (інтерв'ю, кейс-стаді);

– етнографічні спостереження – довготривале спостереження за інтеграцією ШІ у робоче середовище [29].

Зазначені методи є найбільш доречними для рівнів 2–4, де автономність систем зростає, а отже, підвищується потреба в етичній, соціальній та інтерпретаційній оцінці рішень, які приймає ШІ.

3. *Змішані підходи* поєднують кількісні та якісні методи, дозволяють комплексно оцінити довіру:

– тріангуляція даних – комбінація метрик з експертною оцінкою;

– лонгітюдні дослідження – вивчення змін довіри з часом;

– адаптивне тестування – тестові сценарії, які змінюються залежно від рівня автономності.

Такі підходи рекомендовані для рівнів 3–4, де довіра має динамічний і багатогранний характер.

Калібрування довіри – це процес узгодження суб'єктивного рівня довіри дослідника з об'єктивними можливостями системи. Невідповідність між ними призводить до завищеної довіри (over-trust) або заниженої довіри (under-trust), що може мати критичні наслідки для наукової достовірності [16, 25].

Засоби калібрування відповідають за: прозорість обмежень і джерел невизначеності; поступове

нарощування автономності; надання зворотного зв'язку щодо дій системи; адаптацію поведінки ШІ до досвіду користувача.

Довіра до ШІ є динамічним, контекстно-залежним явищем, яке формується на основі досвіду, середовища, особистих рис дослідника та властивостей системи [7, 20]. Тому адаптивні системи повинні враховувати індивідуальні моделі довіри та забезпечувати можливість її динамічного калібрування.

ОБГОВОРЕННЯ ОТРИМАНИХ РЕЗУЛЬТАТІВ

Запропонована таксономія рівнів автономності та профілів довіри має практичне значення для різних аспектів наукової діяльності. Вона може слугувати основою для розробки етичних стандартів застосування ШІ в дослідженнях; створення методик валідації; навчання дослідників роботі з автономними системами; визначення меж відповідальності між людиною та ШІ на різних етапах дослідження.

Дослідникам рекомендується оцінювати рівень автономності систем перед їх використанням; адаптувати методи оцінки довіри відповідно до рівня автономності; враховувати динамічний характер довіри під час довготривалої роботи; забезпечувати прозорість рішень ШІ на всіх етапах дослідницького процесу.

ВИСНОВКИ

Проведене дослідження вирішує актуальну методологічну проблему систематизації рівнів автономності штучного інтелекту в контексті наукових досліджень. Запропонована п'ятирівнева таксономія створює концептуальну основу для класифікації систем ШІ від пасивних інструментів до повністю автономних наукових агентів, враховуючи специфіку дослідницької діяльності, когнітивну складність завдань та етичні вимоги до прозорості.

Розроблені профілі довіри інтегрують когнітивні, афективні та поведінкові компоненти, що дозволяє формувати адекватні очікування щодо можливостей і обмежень систем ШІ на кожному рівні автономності. Систематизація методик валідації довіри забезпечує комплексний підхід до оцінки придатності систем для використання в наукових дослідженнях, сприяючи формуванню каліброваної довіри. Результати дослідження створюють методологічну основу для безпечної інтеграції систем штучного інтелекту у наукову практику, забезпечуючи калібровану довіру, яка відповідає реальним можливостям системи та мінімізує ризики неадекватного делегування відповідальності.

Теоретичне значення дослідження полягає у створенні міждисциплінарної моделі, яка поєднує концепції з галузей штучного інтелекту, психології довіри, когнітивної ергономіки та філософії науки. Практичне значення полягає у формуванні етичних стандартів застосування штучного інтелекту, розробці методик валідації та визначенні меж відповідальності між людиною та ШІ на різних етапах дослідження, що безпосередньо сприяє безпечній інтеграції таких систем у наукову практику.

Обмеження дослідження пов'язані з його теоретичним характером та потребою в емпіричній валідації запропонованої моделі на різних дисциплінах. Культурні та інституційні особливості формування довіри також потребують додаткового вивчення.

Перспективи подальших досліджень можуть стосуватися: емпіричної валідації таксономії на представницьких вибірках дослідників різних дисциплін; розробки автоматизованих інструментів діагностики рівня автономності систем ШІ; дослідження динаміки довіри в довгострокових проєктах з використанням автономних систем; вивчення культурних та дисциплінарних особливостей сприйняття автономності ШІ; створення адаптивних інтерфейсів для калібрування довіри в режимі реального часу; дослідження впливу автономних систем на розвиток наукового мислення та освітні практики; розробку нормативно-правових рамок відповідальності за рішення автономних наукових систем.

REFERENCES

[1] Soder, L., Smakman, Ju., Dunlop, C., & Sussman, O. (2024). An Autonomy-Based Classification: AI Agents, Liability and learnings from the UK Automated Vehicles Act. *INTERFACE Project 2nd Workshop on Regulatable ML at NeurIPS*. Retrieved from https://www.interface-eu.org/publications/ai-agent-classification.

[2] Blab, M., Gimpel, H., & Karnebogen, P. (2024). A Taxonomy and Archetypes of AI-Based Health Care Services: Qualitative Study. *Journal of Medical Internet Research*, 26. https://doi.org/10.2196/53986

[3] Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. *Standards J3016_202104*. Retrieved from https://www.sae.org/standards/content/j3016_202104/

[4] Akbar, S., Coiera. E., & Magrabi, F. (2024). Trustworthy artificial intelligence: A decision-making taxonomy of potential challenges. *Software: Practice and Experience*, 54(9), 1621–1650. https://doi.org/10.1002/spe.3216

[5] Hauptman, A. I., Schelble, B. G., Duan, W., et al. (2024). Understanding the influence of AI autonomy on AI explainability levels in human-AI teams using a mixed methods approach. *Cognition Technology Work*, 26, 435–455. https://doi.org/10.1007/s10111-024-00765-7

[6] Nam, C., Walker, P., Li, H., Lewis, M., & Sycara, K. (2020). Models of Trust in Human Control of Swarms with Varied Levels of Autonomy. *IEEE Transactions on Human-Machine Systems*, 50(3), 194–204. https://doi.org/10.1109/THMS.2019.2896845

[7] Wu, M., Wang, N., & Yuen, K. (2023). Can autonomy level and anthropomorphic characteristics affect public acceptance and trust towards shared autonomous vehicles? *Technological Forecasting and Social Change*, 189. https://doi.org/10.1016/j.techfore.2023.122384

[8] Trofymenko, O. G., Loboda, Yu. G., Hura, V. I., Dyka, A. I., & Strilets, M. I. (2024). Instrumenty shtuchnoho intelektu dlia systemnoho analizu [Artificial intelligence tools for system analysis]. *Visnyk Khersonskoho natsionalnoho tekhnichnoho universytetu – Visnyk of Kherson National Technical University*, 4(91), 349–357. https://doi.org/10.35546/kntu2078-4481.2024.4.46 [in Ukrainian].

[9] Trofymenko, O. G., Loboda, Yu. G., Dyka, A. I., Milchenko, O. O., & Strilets, M. I. (2024). Shtuchnyi intelekt u systemnomu analizi [Artificial intelligence in system analysis]. *Tavriiskyi naukovyi visnyk. Seriia: Tekhnichni nauky – Taurida Scientific Herald. Series: Technical Sciences*, 5, 85–97. https://doi.org/10.32782/tnv-tech.2024.5.9 [in Ukrainian].

[10] Bostrom, A., Hwang, J., et al. (2024). Trust and trustworthy artificial intelligence: A research agenda for AI in the environmental sciences. *Risk Analysis*, 44(6), 1498-1513. https://doi.org/10.1111/risa.14245

[11] Afroogh, S., Akbari, A., Malone, E., et al. (2024). Trust in AI: progress, challenges, and future directions. *Humanities and Social Sciences Communications*, 11. https://doi.org/10.1057/s41599-024-04044-8

[12] Lahusen, C., Maggetti, M., & Slavkovik, M. (2024). Trust, trustworthiness, and AI governance. *Scientific Reports*, 14. https://doi.org/10.1038/s41598-024-71761-0

[13] Kowald, D., Scher, S., et al. (2024). Establishing and evaluating trustworthy AI: overview and research challenges. *Frontiers in Big Data*. Sec. *Machine Learning and Artificial Intelligence*, 7. https://doi.org/10.3389/fdata.2024.1467222

[14] Sadat, N. & Shuttari, M. (2024). Artificial Intelligence for Global Health: Difficulties and Challenges: A Narrative Review. *Open Journal of Epidemiology*, 14, 122–130. https://doi.org/10.4236/ojepi.2024.141009

[15] Hu, Q., Lu, Y., Pan, Z., Gong, Y., Yang, Z. (2021). Can AI artifacts influence human cognition? The effects of artificial autonomy in intelligent personal assistants. *International Journal of Information Management*, 56, 102250. https://doi.org/10.1016/j.ijinfomgt.2020.102250

[16] Abbass, H., Leu, G., & Merrick, K. (2016). A Review of Theoretical and Practical Challenges of Trusted Autonomy in Big Data. *IEEE Access*, 4, 2808–2830. https://doi.org/10.1109/access.2016.2571058

[17] Thiebes, S., Lins, S., & Sunyaev, A. (2020). Trustworthy artificial intelligence. *Electronic Markets*, 31, 447–464. https://doi.org/10.1007/s12525-020-00441-4.

[18] Riedl, R. (2022). Is trust in artificial intelligence systems related to user personality? Review of empirical evidence and future research directions. *Electronic Markets*, 32, 2021–2051. https://doi.org/10.1007/s12525-022-00594-4

[19] Shareef, M., Ahmed, J., Giannakis, M., Dwivedi, Y., Kumar, V., Butt, I., & Kumar, U. (2023). Machine autonomy for rehabilitation of elderly people: A trade-off between machine intelligence and consumer trust. *Journal of Business Research*. 164. https://doi.org/10.1016/j.jbusres.2023.113961

[20] Mcneese, N., Demir, M., Chiou, E., & Cooke, N. (2021). Trust and Team Performance in Human-Autonomy Teaming. *International Journal of Electronic Commerce*, 25, 51–72. https://doi.org/10.1080/10864415.2021.1846854

[21] Li, Y., Wu, B., Huang, Y., & Luan, S. (2024). Developing trustworthy artificial intelligence: insights from research on interpersonal, human-automation, and human-AI trust. *Frontiers in Psychology*, 15. https://doi.org/10.3389/fpsyg.2024.1382693

[22] Darius-Aurel, F., Jacobsen, L., Søndergaard, H., & Otterbring, T. (2023). In companies we trust: consumer adoption of artificial intelligence services and the role of trust in companies and AI autonomy. *Information Technology & People*, 36(8), 155–173. https://doi.org/10.1108/ITP-09-2022-0721

[23] Simas, G. & Ulbricht, V. (2024). Human-AI Interaction: An Analysis of Anthropomorphization and User Engagement in Conversational Agents with a Focus on ChatGPT. *Intelligent Human Systems Integration (IHSI 2024)*, 119. http://doi.org/10.54941/ahfe1004510

[24] Kapoor, A. & Verma, V. (2024). Emotion AI: Understanding emotions through artificial intelligence. *International Journal of Engineering Science and Humanities*, 1(1), 223–232. http://doi.org/10.62904/0vcbvb24

[25] Lukyanenko, R., Maass, W., & Storey, V. (2022). Trust in artificial intelligence: From a Foundational Trust Framework to emerging research opportunities. *Electronic Markets*, 32, 1993–2020. http://doi.org/10.1007/s12525-022-00605-4

[26] Janhunen, E., Toivikko, T., Blomqvist, K., & Siemon, D. (2024). Trust in Digital Human-AI Team Collaboration: A Systematic Review. *Americas' Conference on Information Systems 2024 Proceedings*, 3. Retrieved from https://aisel.aisnet.org/amcis2024/cnow/cnow/3

[27] David, M. W. (2020). Powers Evaluation: from precision, recall and F-measure to ROC, informedness, markedness, and correlation. *International Journal of Machine Learning Technology*. https://arxiv.org/abs/2010.16061

[28] Ding, S., Pan, X., Hu, L., & Liu, L. (2025). A new model for calculating human trust behavior during human-AI collaboration in multiple decision-making tasks: A Bayesian approach. *Computers & Industrial Engineering*, 200. https://doi.org/10.1016/j.cie.2025.110872

[29] Vereschak, O., Bailly, G., & Caramiaux, B. (2021). How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies. <i>Proceedings of the ACM on Human-Computer Interaction<\i>, l(5), CSCW2, 1-39. https://doi.org/10.1145/3476068

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

[1] Soder, L., Smakman, Ju, Dunlop, C., & Sussman, O. (2024). An Autonomy-Based Classification: AI Agents, Liability and learnings from the UK Automated Vehicles Act. *INTERFACE Project 2nd Workshop on Regulatable ML at NeurIPS*. URL: https://www.interface-eu.org/publications/ai-agent-classification

[2] Blab, M., Gimpel, H., Karnebogen, P. (2024). A Taxonomy and Archetypes of AI-Based Health Care Services: Qualitative Study. *Journal of Medical Internet Research*, 26. https://doi.org/10.2196/53986

[3] Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. *Standards J3016_202104*. URL: https://www.sae.org/standards/content/j3016_202104/

[4] Akbar, S., Coiera. E., & Magrabi, F. (2024). Trustworthy artificial intelligence: A decision-making taxonomy of potential challenges. *Software: Practice and Experience*, 54(9), 1621–1650. https://doi.org/10.1002/spe.3216

[5] Hauptman, A. I., Schelble, B. G., Duan, W., et al. (2024). Understanding the influence of AI autonomy on AI explainability levels in human-AI teams using a mixed methods approach. *Cognition Technology Work*, 26, 435–455. https://doi.org/10.1007/s10111-024-00765-7

[6] Nam, C., Walker, P., Li, H., Lewis, M., & Sycara, K. (2020). Models of Trust in Human Control of Swarms with Varied Levels of Autonomy. *IEEE Transactions on Human-Machine Systems*, 50(3), 194–204. https://doi.org/10.1109/THMS.2019.2896845

[7] Wu, M., Wang, N., & Yuen, K. (2023). Can autonomy level and anthropomorphic characteristics affect public acceptance and trust towards shared autonomous vehicles? *Technological Forecasting and Social Change*, 189. https://doi.org/10.1016/j.techfore.2023.122384

[8] Trofymenko, O. G., Loboda, Yu. G., Hura, V. I., Dyka, A. I., & Strilets, M. I. Artificial intelligence tools for system analysis. *Visnyk of Kherson National Technical University*. 2024, 4(91), 349–357. DOI: https://doi.org/10.35546/kntu2078-4481.2024.4.46

[9] Trofymenko, O. G., Loboda, Yu. G., Dyka, A. I., Milchenko, O. O., & Strilets, M. I. (2024). Artificial intelligence in system analysis. *Taurida Scientific Herald. Series: Technical Sciences*, 5, 85–97. https://doi.org/10.32782/tnv-tech.2024.5.9

[10] Bostrom, A., Hwang, J., et al. (2024). Trust and trustworthy artificial intelligence: A research agenda for AI in the environmental sciences. *Risk Analysis*, 44(6), 1498–1513. https://doi.org/10.1111/risa.14245

[11] Afroogh, S., Akbari, A., Malone, E., et al. (2024). Trust in AI: progress, challenges, and future directions. *Humanities and Social Sciences Communications*, 11. https://doi.org/10.1057/s41599-024-04044-8

[12] Lahusen, C., Maggetti, M., & Slavkovik, M. (2024). Trust, trustworthiness and AI governance. *Scientific Reports*, 14. https://doi.org/10.1038/s41598-024-71761-0

[13] Kowald, D., Scher, S., et al. (2024). Establishing and evaluating trustworthy AI: overview and research challenges. *Frontiers in Big Data*. Sec. *Machine Learning and Artificial Intelligence*, 7. https://doi.org/10.3389/fdata.2024.1467222

[14] Sadat, N. & Shuttari, M. (2024). Artificial Intelligence for Global Health: Difficulties and Challenges: A Narrative Review. *Open Journal of Epidemiology*, 14, 122–130. https://doi.org/10.4236/ojepi.2024.141009

[15] Hu, Q., Lu, Y., Pan, Z., Gong, Y., Yang, Z. (2021). Can AI artifacts influence human cognition? The effects of artificial autonomy in intelligent personal assistants. *International Journal of Information Management*, 56, 102250. https://doi.org/10.1016/j.ijinfomgt.2020.102250

[16] Abbass, H., Leu, G., & Merrick, K. (2016). A Review of Theoretical and Practical Challenges of Trusted Autonomy in Big Data. *IEEE Access*, 4, 2808-2830. https://doi.org/10.1109/access.2016.2571058

[17] Thiebes, S., Lins, S., & Sunyaev, A. (2020). Trustworthy artificial intelligence. *Electronic Markets*, 31, 447–464. https://doi.org/10.1007/s12525-020-00441-4

[18] Riedl, R. (2022). Is trust in artificial intelligence systems related to user personality? Review of empirical evidence and future research directions. *Electronic Markets*, 32, 2021–2051. https://doi.org/10.1007/s12525-022-00594-4

[19] Shareef, M., Ahmed, J., Giannakis, M., Dwivedi, Y., Kumar, V., Butt, I., & Kumar, U. (2023). Machine autonomy for rehabilitation of elderly people: A trade-off between machine intelligence and consumer trust. *Journal of Business Research*. 164. https://doi.org/10.1016/j.jbusres.2023.113961

[20] Mcneese, N., Demir, M., Chiou, E., & Cooke, N. (2021). Trust and Team Performance in Human-Autonomy Teaming. *International Journal of Electronic Commerce*, 25, 51–72. https://doi.org/10.1080/10864415.2021.1846854

[21] Li, Y., Wu, B., Huang, Y., & Luan, S. (2024). Developing trustworthy artificial intelligence: insights from research on interpersonal, human-automation, and human-AI trust. *Frontiers in Psychology*, 15. https://doi.org/10.3389/fpsyg.2024.1382693

[22] Darius-Aurel, F., Jacobsen, L., Søndergaard, H., & Otterbring, T. (2023). In companies we trust: consumer adoption of artificial intelligence services and the role of trust in companies and AI autonomy. *Information Technology & People*, 36(8), 155–173. https://doi.org/10.1108/ITP-09-2022-0721

[23] Simas, G. & Ulbricht, V. (2024). Human-AI Interaction: An Analysis of Anthropomorphization and User Engagement in Conversational Agents with a Focus on ChatGPT. *Intelligent Human Systems Integration (IHSI 2024)*, 119. http://doi.org/10.54941/ahfe1004510

[24] Kapoor, A. & Verma, V. (2024). Emotion AI: Understanding emotions through artificial intelligence. *International Journal of Engineering Science and Humanities*, 1(1), 223–232. http://doi.org/10.62904/0vcbvb24

[25] Lukyanenko, R., Maass, W., & Storey, V. (2022). Trust in artificial intelligence: From a Foundational Trust Framework to emerging research opportunities. *Electronic Markets*, 32, 1993–2020. http://doi.org/10.1007/s12525-022-00605-4

[26] Janhunen, E., Toivikko, T., Blomqvist, K., & Siemon, D. (2024). Trust in Digital Human-AI Team Collaboration: A Systematic Review. *Americas' Conference on Information Systems 2024 Proceedings*, 3. URL: https://aisel.aisnet.org/amcis2024/cnow/cnow/3

[27] David, M. W. (2020). Powers Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *International Journal of Machine Learning Technology*. URL: https://arxiv.org/abs/2010.16061

[28] Ding, S., Pan, X., Hu, L., & Liu, L. (2025). A new model for calculating human trust behavior during human-AI collaboration in multiple decision-making tasks: A Bayesian approach. *Computers & Industrial Engineering*, 200, https://doi.org/10.1016/j.cie.2025.110872

[29] Vereschak, O., Bailly, G., & Caramiaux, B. (2021). How to Evaluate Trust in AI-Assisted Decision Making? A Survey of Empirical Methodologies. *Proceedings of the ACM on Human-Computer Interaction*, l(5), CSCW2, 1–39. https://doi.org/10.1145/3476068

Дата надходження статті до редакції: 28.07.2025
Дата затвердження статті до друку: 13.08.2025
Опубліковано: 24.11.2025