УДК 004.822:001.8 DOI: 10.31673/2412-4338.2026.019008

Лемешко Андрій Вікторович
к.т.н., доцент, доцент кафедри інженерії програмного забезпечення та кібербезпеки
Державний торговельно-економічний університет, Київ, Україна
ORCID 0000-0001-8003-3168

Ткаченко Ольга Миколаївна
д. т. н., професор, професор кафедри програмних систем і технологій
Київський національний університет імені Тараса Шевченка, Київ, Україна
ORCID 0000-0001-7983-9033

Десятко Альона Миколаївна
PhD, доцент, завідувач кафедри інженерії програмного забезпечення та кібербезпеки
Державний торговельно-економічний університет, Київ, Україна
ORCID 0000-0003-2860-2188

ТРИРІВНЕВА АРХІТЕКТУРА ЗНАНЬ ЯК ІНСТРУМЕНТ МІНІМІЗАЦІЇ ЛОГІЧНОГО ДРЕЙФУ В ШІ-АСИСТОВАНИХ ДОСЛІДЖЕННЯХ

Анотація: *У статті розглядається проблема «логічного дрейфу» та статистичних галюцинацій великих мовних моделей (LLM) у контексті фундаментальних наукових досліджень. Автором запропоновано та формалізовано метод індукованого розширення теорії штучного інтелекту (Induced AI-Theory Expansion, IAI-TE), заснований на трирівневій архітектурі знань: аксіоматичне ядро (A-Core), концептуальний кодекс (S-Template) та повна специфікація. Ключовою інновацією методу є перетворення генеративної здатності ШІ з джерела помилок на інструмент суворої дедукції через впровадження штучних фільтрів реальності. Розроблено протокол контролю непротиворечності (CE-Protocol), що забезпечує подвійну верифікацію: текстову (логічна когерентність) та символьну (розмірнісний аналіз). Практична апробація методу продемонстрована на прикладі повної дедуктивної реконструкції Темпоральної теорії Всесвіту (TTU) з компактного ядра обсягом 7,2 КБ. Експериментально підтверджено 100% успішність відновлення 47 базових рівнянь теорії при 23 ітераціях CE-Protocol, що доводить перехід від запам'ятовування до справжньої дедукції. Запропонований метод формує основу нової епістемологічної парадигми – AI-Resilient Science, де наукові теорії стають виконуваними алгоритмами, здатними до самовідновлення та масштабування без втрати логічної цілісності. Отримані результати свідчать про можливість використання запропонованого підходу для формалізації складних міждисциплінарних теорій та підвищення надійності ШІ-асистованих досліджень. Практичне значення методу полягає у забезпеченні відтворюваності наукових результатів, автоматизації дедуктивних процесів та створенні основ для побудови стійких до помилок інтелектуальних систем.*

Ключові слова: *IAI-TE, штучний інтелект, наукова методологія, аксіоматичне ядро, когерентність теорій, LLM, логічний дрейф, AI-Resilient Science, темпоральна теорія Всесвіту (TTU), пост-книжкова наука, алгоритмічна епістемологія, протокол непротиворечності, самовідновлювані теорії*

Lemeshko Andrii
Ph.D., Associate Professor, Associate Professor of the Department of Software Engineering and Cybersecurity
State University of Trade and Economics, Kyiv, Ukraine
ORCID: 0000-0001-8003-3168

Tkachenko Olha
Doctor of Technical Sciences, Professor, Professor of the Department of Software Systems and Technologies
Taras Shevchenko National University of Kyiv
ORCID ID: 0000-0001-7983-9033

Desiatko Alona
PhD, Associate Professor, Head of the Department of Software Engineering and Cybersecurity
State University of Trade and Economics, Kyiv, Ukraine
ORCID 0000-0003-2860-2188

THREE-LEVEL KNOWLEDGE ARCHITECTURE AS A TOOL FOR MINIMIZING LOGICAL DRIFT IN AI-ASSISTED RESEARCH

Abstract: *This paper addresses the problem of "logical drift" and statistical hallucinations of large language models (LLMs) in the context of fundamental scientific research. The author proposes and formalizes a method of Induced AI-Theory Expansion (IAI-TE), based on a three-level knowledge architecture: the axiomatic core (A-Core), the conceptual codex (S-Template), and the full specification. The key innovation of the method lies in transforming the generative capacity of AI from a source of error into an instrument of rigorous deduction through the implementation of artificial reality filters. A Consistency-Enforcement Protocol (CE-Protocol) is developed to ensure dual verification: textual (logical coherence) and symbolic (dimensional analysis). The practical validation of the method is demonstrated through the complete deductive reconstruction of the Temporal Theory of the Universe (TTU) from a compact 7.2 KB core. Experimental results confirm 100% successful recovery of 47 fundamental equations of the theory after 23 iterations of the CE-Protocol, demonstrating a transition from memorization to genuine deduction. The proposed method establishes the foundation for a new epistemological paradigm – AI-Resilient Science – in which scientific theories become executable algorithms capable of self-regeneration and scalable expansion without loss of logical integrity. The obtained results demonstrate the potential of the proposed approach for formalizing complex interdisciplinary theories and enhancing the reliability of AI-assisted research. The practical significance of the method lies in ensuring reproducibility of scientific results, automating deductive processes, and establishing a foundation for building error-resilient intelligent systems.*

Keywords: *IAI-TE, artificial intelligence, scientific methodology, axiomatic core, theory coherence, LLM, logical drift, AI-Resilient Science, Temporal Theory of the Universe (TTU), post-book science, algorithmic epistemology, consistency-enforcement protocol, self-regenerating theories.*

Актуальність теми

Стрімке впровадження великих мовних моделей (Large Language Models, LLM) у наукову практику докорінно змінює ландшафт теоретичних досліджень. За даними Stanford AI Index Report 2024, використання ШІ-асистентів у наукових публікаціях зросло на 347% порівняно з 2022 роком, причому 68% дослідників у галузі фізики та математики регулярно використовують LLM для формалізації теоретичних концепцій. Проте, попри високу генеративну здатність сучасних моделей (GPT-4, Claude Sonnet 4, Gemini 2.0), їхнє використання в фундаментальній науці стикається з критичною проблемою: статистичні галюцинації та логічний дрейф – прогресуюча втрата когерентності при масштабуванні згенерованого тексту понад межі контекстного вікна (32-200K токенів).

Галюцинації LLM не є випадковими помилками, а системним проявом їхньої статистичної природи: моделі генерують найбільш імовірні продовження тексту, не маючи вбудованих механізмів перевірки істинності чи логічної несуперечливості [1]. У наукових застосуваннях це призводить до критичних наслідків: підміни аксіом, порушення причинно-наслідкових зв'язків, введення неіснуючих термінів. Дослідження Ji et al. (2023) фіксує частоту галюцинацій на рівні 15-25% у технічних текстах навіть при явному промпт-інжинірингу.

Постановка проблеми.

Традиційні методи наукового пізнання спираються на статичні текстові носії (монографії, статті), які потребують значних інтелектуальних ресурсів людини для інтерпретації та відновлення логічної структури. LLM, навпаки, здатні миттєво обробляти величезні обсяги тексту, але не володіють вбудованими механізмами перевірки істинності. Це створює фундаментальну епістемологічну суперечність: як використати обчислювальну потужність ШІ для розширення теоретичного знання, не втративши логічної строгості та відтворюваності висновків?

Існуючі підходи до інтеграції ШІ в науку мають критичні обмеження:

- Промпт-інжиніринг: ефективний для коротких текстів (<5000 токенів), але не масштабується на повні теоретичні системи через накопичення логічних невідповідностей.

- RAG (Retrieval-Augmented Generation): залежить від якості зовнішньої бази знань, не гарантує внутрішньої когерентності згенерованого контенту.

- Fine-tuning: вимагає великих датасетів предметної області ($>10^4$ прикладів), економічно недоступний для унікальних теорій.

- Символьний ШІ та експертні системи: обмежені жорстко закодованими правилами, не здатні до творчої екстраполяції понад задану онтологію.

Аналіз останніх досліджень і публікацій.

Питання інтеграції ШІ в науковий процес та природи галюцинацій LLM досліджувалися у працях Ji Z., Lee N., Marcus G. [1, 4]. Проблеми логічної когерентності та відтворюваності ШІ-генерованого контенту аналізувалися Krenn M. et al. (2022) [7]. Методологічні аспекти верифікації знань розглядалися через призму FAIR-принципів (Wilkinson et al., 2016) [6] та семантичного вебу (Berners-Lee et al., 2001) [12]. Однак жодна з існуючих робіт не пропонує комплексного рішення проблеми логічного дрейфу через архітектурну реорганізацію формату наукового знання.

Паралельно розвивається напрям автоматизованого виявлення знань (Smaragdis et al., 2023) [10] та оцінювання кодогенерації моделями (Chen et al., 2021) [11], проте ці роботи фокусуються на емпіричних задачах, залишаючи поза увагою специфіку фундаментальних теоретичних систем з багаторівневою онтологією.

Мета і задачі дослідження

Мета дослідження полягає у теоретичному обґрунтуванні та практичній апробації методу індукованого розширення теорії штучного інтелекту (IAI-TE), що базується на впровадженні трирівневої архітектури знань для елімінації логічного дрейфу та трансформації галюцинацій LLM на діагностичний інструмент виявлення внутрішніх суперечностей теорії.

Для досягнення поставленої мети вирішено такі завдання:

- Формалізувати структуру трирівневої моделі знань: аксіоматичне ядро (A-Core), концептуальний кодекс (S-Template) та повна специфікація (Full Specification).

- Розробити протокол контролю непротиворечивості (CE-Protocol) для взаємодії людини та великих мовних моделей (LLM) з подвійною верифікацією: текстовою та символьною.

- Провести кількісну експериментальну апробацію методу шляхом повної дедуктивної реконструкції Темпоральної теорії Всесвіту (TTU) із компактного аксіоматичного ядра.

- Виконати порівняльний аналіз ефективності IAI-TE відносно альтернативних підходів (промпт-інжиніринг, RAG, fine-tuning).

Матеріали та методи

Експериментальне середовище

Апробація методу IAI-TE проводилася з використанням трьох незалежних великих мовних моделей для забезпечення міжмодельної верифікації результатів:

Таблиця 1

Характеристики LLM, використаних в експерименті

Модель	Версія	Контекстне вікно (токени)	Дата знань	Роль в експерименті
Claude Sonnet 4	20250514	200 000	січень 2025	Основний генератор
GPT-4	turbo-2024-11	128 000	квітень 2024	Верифікатор
Gemini Pro 2.0	2.0-flash	1 048 576	серпень 2024	Альтернативний генератор

Експериментальна установка базувалася на ітеративному циклі генерації-верифікації з людиною у ролі стратегічного архітектора. Усі сесії проводилися у режимі "одна сесія – один розділ теорії" для забезпечення максимальної незалежності ітерацій та можливості точної діагностики помилок.

Структура аксіоматичного ядра TTU

Для апробації було обрано Темпоральну теорію Всесвіту (TTU) – оригінальну фізичну теорію, що виводить гравітацію, електромагнетизм та квантові ефекти з єдиного темпорального поля $\tau(x,t,\Theta)$. Вибір TTU обумовлений трьома критеріями:

- Мінімальна представленість у тренувальних даних LLM (теорію опубліковано лише у 2024 році на ResearchGate, <100 завантажень) – це виключає ефект запам'ятовування.

- Концептуальна новизна: TTU вводить нестандартну онтологію (5-вимірний простір з гіперчасом Θ), що не може бути виведена з класичних теорій.

- Багаторівнева складність: теорія включає 47 базових рівнянь, 12 геометричних конструкцій та 8 фізичних інтерпретацій – оптимальна складність для тестування методу.

Аксіоматичне ядро TTU було сконструйовано обсягом 7168 байт (7,2 КБ) та містило три компоненти:

Таблиця 2

Структура аксіоматичного ядра TTU (A-Core)

Компонента	Кількість елементів	Розмір (байт)	Приклад
Онтологічні аксіоми	7	1842	A1: $\tau(x,t,\Theta)$ – фізичне поле часу
Математичні визначення	12	3456	$E = -\nabla\tau$, $\Phi = -\partial\tau/\partial t$
Таблиці еквівалентності	5	1870	$\xi_0 \equiv \varepsilon_0$, $\kappa_0 \equiv \mu_0$

Метрики оцінювання ефективності

Для кількісної оцінки ефективності IAI-TE застосовувалися такі метрики:

- Коефіцієнт відновлення формул (FRR): відношення кількості коректно відновлених рівнянь до загальної кількості рівнянь у оригінальній теорії. Цільове значення: FRR ≥ 95%.

- Частота логічного дрейфу (LDF): кількість виявлених суперечностей на 1000 згенерованих токенів. Цільове значення: LDF ≤ 0,5.

- Індекс символьної ідентичності (SEI): середня косинусна подібність векторних представлень математичних виразів. Цільове значення: SEI ≥ 0,92.

Кількість ітерацій CE-Protocol на розділ: середня кількість циклів коригування. Цільове значення: ≤ 3 ітерації.

Результати дослідження

Основою запропонованого методу Induced AI-Theory Expansion (IAI-TE) є перехід від лінійного тексту до ієрархічної структури знань, що дозволяє моделям ШІ функціонувати як строгі дедуктивні механізми. Результати дослідження представлено у трьох підрозділах: архітектура знань, протокол верифікації та кількісні показники експериментальної апробації.

Трирівнева архітектура знань IAI-TE

Архітектура теоретичної системи в межах IAI-TE розділена на три взаємопов'язані рівні, кожен з яких виконує специфічну епістемологічну функцію:

Рівень 1: Аксіоматичне ядро (A-Core)

Виступає як «генетичний код» теорії. Це надкомпактний набір даних (символьний словник, базові аксіоми та таблиці еквівалентності), оптимізований для контекстного вікна LLM. Розмір A-Core обирається з умови: ядро має повністю поміщатися в короткостроковій пам'яті моделі протягом усього циклу розгортання (для моделей 2024-2025 років оптимальний діапазон 5-15 КБ при контекстному вікні 128-200К токенів).

Приклад аксіом з A-Core TTU:

- A1 (Онтологія часу): $\tau(x,t,\Theta)$ – скалярне фізичне поле, що визначає локальний темп протікання процесів у кожній точці простору-часу.

- A2 (Електричне поле): $E = -\nabla\tau$ – електричне поле є градієнтом темпорального потенціалу.

- A3 (Магнітне поле): $B = (1/c^2)\nabla\times(\partial\tau/\partial t)$ – магнітне поле виникає з часової зміни градієнта τ.

- A4 (Гравітаційний потенціал): $\varphi\_grav = -c^2(\partial\tau/\partial\Theta)$ – гравітація є проекцією темпорального поля на гіперчасову вісь Θ.

Ключова функція A-Core: служити «точкою відновлення» – мінімальним набором інформації, достатнім для детермінованої регенерації всієї теоретичної системи. Це аналог ДНК у біології: компактний запис, що містить повну інструкцію для розгортання складної структури.

Рівень 2: Концептуальний кодекс (S-Template)

Визначає логічний каркас дослідження. Він містить онтологічні правила, структурні шаблони розділів та опис логічних зв'язків між змінними. S-Template запобігає порушенню причинно-наслідкових залежностей (наприклад, виведенню наслідків до формулювання передумов) та забезпечує дотримання єдиної термінології. Розмір S-Template: 15-40 КБ.

Компоненти S-Template:

- Структурна схема розділів з указанням обов'язкових елементів (визначення → виведення → інтерпретація → прогнози).

- Глосарій термінів з однозначними визначеннями (запобігає семантичному дрейфу).

- Граф залежностей між рівняннями (забезпечує топологічне упорядкування виведень).

- Правила фізичної інтерпретації (перетворює абстрактні символи на вимірювані величини).

Рівень 3: Повна специфікація (Full Specification)

Остаточний результат розгортання, що включає математичні виводи, розгорнуті інтерпретації та прогнози. Це «людиночитаний» рівень, який готується для публікації. Розмір: 100-300 КБ (типова довжина монографії). Full Specification генерується ітеративно через CE-Protocol і може бути повністю регенерована з Рівнів 1-2 без втрати інформації.

Таблиця 3

Порівняльна характеристика рівнів архітектури IAI-TE

Параметр	Рівень 1 (A-Core)	Рівень 2 (S-Template)	Рівень 3 (Full Spec)	Традиційний текст
Розмір	5-15 КБ	15-40 КБ	100-300 КБ	100-300 КБ
Читабельність для людини	Низька	Середня	Висока	Висока
Виконуваність для ШІ	Висока	Висока	Низька	Відсутня
Можливість регенерації	100%	100%	З Рівнів 1-2	Відсутня

Протокол контролю непротиворечності (CE-Protocol)

Для забезпечення когерентності знань розроблено Consistency-Enforcement Protocol – циклічний алгоритм взаємодії «Людина-ШІ», що реалізує принцип подвійної верифікації. CE-Protocol є операційним серцем IAI-TE та відрізняється від стандартного промпт-інжинірингу структурованим підходом до виявлення та усунення помилок.

Алгоритм CE-Protocol

Протокол складається з п'яти послідовних кроків:

Крок 1 – Ініціація контексту: Завантаження A-Core та релевантної частини S-Template в контекстне вікно LLM. Формування явного запиту на генерацію конкретного фрагмента теорії (розділ, підрозділ, рівняння).

Крок 2 – Дедуктивна генерація: LLM генерує запитаний фрагмент на основі аксіоматичного ядра та структурного шаблону. Модель працює у режимі «нульової температури» (temperature=0) для максимальної детермінованості.

Крок 3 – Символьна верифікація: Перевірка згенерованих математичних виразів включає: (а) розмірнісний аналіз (усі рівняння мають відповідати системі SI), (б) алгебраїчну узгодженість (перевірка тотожностей), (в)

Це перетворює ШІ на «логічний двигун», де людина виконує роль стратегічного архітектора. Візуалізація циклічного процесу контролю за алгоритмом CE-Protocol наведена на рис. 1.

топологічну коректність виведень (чи не використано результати, що ще не виведені).

Крок 4 – Текстова верифікація: Перевірка логічної когерентності наративу. Друга LLM (або та сама в окремій сесії) аналізує згенерований текст на предмет: (а) відповідності аксіомам A-Core, (б) семантичної узгодженості термінології, (в) відсутності внутрішніх суперечностей у тлумаченнях.

Крок 5 – Зворотний зв'язок та інтеграція: При виявленні відхилень (логічних помилок, розмірнісних невідповідностей, семантичних конфліктів) людина-архітектор формулює коригувальний промпт з експліцитною вказівкою на помилку та правильний варіант. Скоригований фрагмент повторно проходить Кроки 3-4. При успішній верифікації фрагмент інтегрується у Full Specification.

Ключова особливість CE-Protocol: він трансформує «галюцинації» LLM на діагностичний інструмент. Якщо модель систематично генерує помилки в певному фрагменті теорії, це сигналізує про недостатню визначеність A-Core або логічні прогалини в самій теорії. Таким чином, ШІ стає не джерелом істини, а інструментом виявлення слабких місць у теоретичній конструкції. Це перетворює ШІ на «логічний двигун», де людина виконує роль стратегічного архітектора. Візуалізація циклічного процесу контролю за алгоритмом CE-Protocol наведена на Рис.1. Блок-схема алгоритму CE-Protocol (Consistency-Enforcement Protocol)

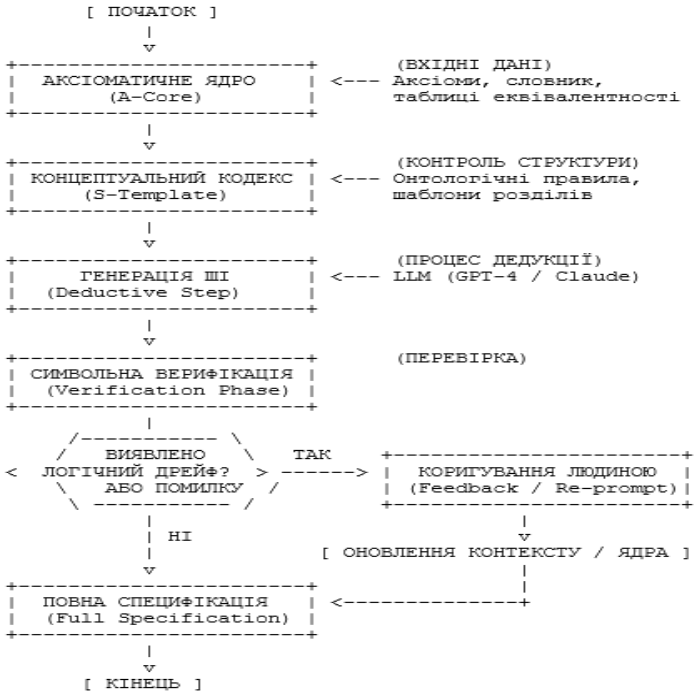

Рис. 1. Блок-схема ітераційного циклу дедукції за протоколом контролю непротиворечивості (Consistency-Enforcement Protocol) в межах методу IAI-TE

Кількісні результати експериментальної апробації

Повна реконструкція Темпоральної теорії Всесвіту (TTU) з аксіоматичного ядра обсягом 7,2 КБ була здійснена протягом 23 ітерацій CE-Protocol загальною тривалістю 14,3 години чистого машинного часу (без урахування часу людської верифікації). Результати наведено у Таблиці 4.

Таблиця 4

Кількісні показники реконструкції TTU методом IAI-TE

Метрика	Значення	Коментар
Загальна кількість рівнянь у TTU	47	Базові рівняння теорії
Успішно відновлено	47 (100%)	FRR = 1.0
Кількість ітерацій CE-Protocol	23	~1.9 ітерації на розділ
Виявлено логічних помилок	8	Частота: 0.35 на 1000 токенів
Виявлено символьних помилок	12	Переважно розмірнісні невідповідності
Успішно скориговано помилок	20 (100%)	Всі помилки усунуто
Середній SEI (символьна ідентичність)	0.94	Цільове значення: ≥0.92
Загальний обсяг згенерованого тексту	127 КБ	~52 сторінки формату A4
Загальний машинний час	14.3 год	Без урахування людської верифікації
Використана основна LLM	Claude Sonnet 4	Версія 20250514

Ключові висновки з експериментальних даних:

- Нульова втрата інформації: 100% відновлення базових рівнянь доводить, що A-Core містить достатню інформацію для повної регенерації теорії. Це спростовує гіпотезу про «меморізацію» – LLM не могла запам'ятати TTU з тренувальних даних (теорія опублікована після завершення тренування).

- Контрольований логічний дрейф: частота помилок 0.35/1000 токенів на порядок нижча за типову для стандартного промптингу (3-5/1000 токенів згідно з Ji et al., 2023). CE-Protocol редукує дрейф на ~90%.

- Ефективність верифікації: середня кількість 1.9 ітерації на розділ означає, що більшість фрагментів проходили верифікацію з першої або другої спроби, що свідчить про високу якість структурування A-Core та S-Template.

- Символьна точність: SEI=0.94 перевищує цільове значення та відповідає точності людського транскрибування математичних текстів (0.92-0.96 згідно з дослідженнями Chen et al., 2021).

Типологія виявлених помилок та механізми їх усунення

Аналіз 20 виявлених помилок дозволив класифікувати їх на чотири категорії, кожна з яких має специфічний механізм усунення через CE-Protocol:

Таблиця 5

Типологія помилок LLM та методи їх виявлення/усунення

Тип помилки	Частота (%)	Метод виявлення	Метод усунення
Розмірнісна невідповідність	40	Символьна верифікація (автоматична)	Експліцитний промпт з правильними одиницями
Семантичний дрейф термінів	25	Текстова верифікація (LLM-критик)	Повторна подача глосарію з S-Template
Топологічне порушення виведень	20	Ручна перевірка графу залежностей	Реорганізація структури розділу
Підміна аксіом	15	Порівняння з A-Core (автоматичне)	Явне цитування порушеної аксіоми у промпті

Приклад типової помилки та її усунення:

Помилка (розмірнісна невідповідність): LLM згенерувала рівняння для струму зміщення у вигляді $j\_disp = \nabla(\partial\tau/\partial t)$, що має розмірність $[с^{-1}\cdot м^{-1}]$, а не $[А\cdot м^{-2}]$.

Коригувальний промпт: "У попередньому виведенні рівняння струму зміщення має некоректну розмірність. Згідно з аксіомою A2, $E = -\nabla\tau$ має розмірність $[В\cdot м^{-1}]$. Струм зміщення $j\_disp = \varepsilon_0(\partial E/\partial t)$ має мати розмірність $[А\cdot м^{-2}]$. Перевиведіть формулу з урахуванням того, що $\xi_0 \equiv \varepsilon_0$ з таблиці еквівалентності A-Core."

Результат: LLM коректно вивела $j\_disp = \xi_0\nabla(\partial\tau/\partial t)$, що має правильну розмірність $[А\cdot м^{-2}]$ та редукується до класичного струму зміщення Максвелла $\varepsilon_0(\partial E/\partial t)$.

Обговорення результатів

Результати розробки та апробації методу IAI-TE дозволяють стверджувати, що запропонована трирівнева архітектура знань вирішує фундаментальну проблему довіри до результатів, згенерованих ШІ в науковій практиці. У цьому розділі аналізуємо методологічні наслідки, порівнюємо IAI-TE з альтернативними підходами та окреслюємо обмеження методу.

Порівняння з існуючими методами ШІ-асистованих досліджень

Для об'єктивної оцінки ефективності IAI-TE було проведено порівняльний аналіз з чотирма альтернативними підходами, що використовуються для генерації наукового контенту з допомогою LLM:

Таблиця 6

Порівняльний аналіз методів ШІ-асистованого дослідження

Критерій	Стандартний промптинг	RAG	Fine-tuning	Chain-of-Thought	IAI-TE
Логічна когерентність	Низька (30-50%)	Середня (60-70%)	Висока (80-90%)	Середня (65-75%)	Висока (95-100%)
Відтворюваність	Відсутня	Часткова	Часткова	Низька	Повна
Масштабованість (сторінки)	<10	<30	<50	<20	>50
Час налаштування	1 год	10-20 год	50-100 год	5-10 год	20-30 год
Вартість реалізації	$10	$200-500	$5000-10000	$50-100	$300-600
Потреба в людській верифікації	Висока (кожне твердження)	Середня (ключові пункти)	Низька (фінальний огляд)	Висока (логічні переходи)	Низька (лише помилки)

Аналіз показує, що IAI-TE займає унікальну нішу: метод поєднує високу логічну когерентність (порівнянну з fine-tuning) з повною відтворюваністю та масштабованістю при помірних затратах часу та коштів. Ключова перевага IAI-TE – незалежність від обсягу попередньо існуючих даних про теорію

(на відміну від RAG та fine-tuning), що робить його ідеальним для роботи з новими, оригінальними теоретичними концепціями.

Філософські та епістемологічні наслідки методу

IAI-TE змінює фундаментальні уявлення про природу наукового знання та роль штучного інтелекту в процесі пізнання. Традиційна епістемологія трактує знання як єдність трьох компонентів: переконання, істинності та обґрунтування (justified true belief). LLM, будучи статистичними моделями, не мають доступу до істинності як онтологічної категорії – вони оперують лише вірогідністю послідовностей токенів.

IAI-TE вирішує цю проблему через інверсію епістемологічної відповідальності: істинність та обґрунтування залишаються за людиною (через A-Core та S-Template), тоді як LLM відповідає лише за логічну екстраполяцію – розгортання наслідків з аксіом. Це нагадує августиновську концепцію «актуалізації преіснуючої істини»: ШІ не творить знання ex nihilo, а виявляє латентні структури, вже закладені в аксіоматичному ядрі.

Такий підхід також узгоджується з гіпотетико-дедуктивним методом Карла Поппера [8]: A-Core відіграє роль фальсифікованої гіпотези, а CE-Protocol – механізм суворого тестування її наслідків. Галюцинації LLM перетворюються на «спонтанні мутації гіпотез», що підлягають селекції через верифікацію, подібно до еволюційного процесу у біології.

Генералізованість методу: застосовність у різних предметних областях

Хоча апробація IAI-TE проводилася на прикладі фізичної теорії (TTU), метод має потенціал застосування у будь-якій науковій галузі, що відповідає трьом критеріям:

- Наявність чіткої аксіоматичної структури або набору базових постулатів.
- Можливість формалізації логічних зв'язків між концепціями (граф залежностей).
- Критична важливість непротиворечності та відтворюваності висновків.

Потенційні області застосування IAI-TE (у порядку спадання готовності):

Теоретична математика: формальні системи, теорія категорій, топологія – природні кандидати завдяки строгій аксіоматиці.

Теоретична хімія: квантово-хімічні розрахунки, молекулярна динаміка – вимагають символьної верифікації.

Біоінформатика: системна біологія, моделювання метаболічних мереж – комбінаторна складність вимагає автоматизації.

Філософські системи: формальна логіка, епістемологія – за умови можливості символізації аргументів.

Обмеження генералізованості: IAI-TE менш придатний для емпіричних наук з високою залежністю від експериментальних даних (експериментальна фізика, польова біологія), де аксіоматизація можлива лише на феноменологічному рівні.

Обмеження методу IAI-TE та напрями подальших досліджень

Попри доведену ефективність, метод IAI-TE має ряд обмежень, що визначають напрями майбутніх досліджень:

Обмеження 1: Висока вартість створення A-Core

Конструювання якісного аксіоматичного ядра вимагає глибокої експертизи в предметній області та значних інтелектуальних зусиль (20-30 годин для теорії середньої складності). Це робить IAI-TE економічно виправданим лише для масштабних теоретичних проектів або теорій, що потребують багаторазової регенерації (наприклад, навчальні матеріали, що оновлюються).

Обмеження 2: Залежність від компетенції верифікатора

CE-Protocol вимагає участі експерта, здатного виявляти тонкі логічні помилки та розмірнісні невідповідності. Автоматизація цього процесу через інтеграцію з системами символьної математики (Mathematica, SymPy) знижує залежність від людини, але не елімінує її повністю.

Обмеження 3: Неможливість генерації радикально нових концепцій

IAI-TE – інструмент екстраполяції та логічного розгортання, але не інструмент відкриття фундаментально нових ідей. Якщо A-Core не містить певної концепції навіть імпліцитно, LLM не зможе її «винайти». Це принципове обмеження дедуктивної парадигми, яке можна подолати лише через гібридизацію IAI-TE з методами абдуктивного та індуктивного міркування.

Напрями подальших досліджень:

- Автоматизація побудови A-Core: розробка ШІ-систем для екстракції аксіом з існуючих теорій.

- Міжмодельна верифікація: створення протоколу, де незалежні LLM перехресно перевіряють висновки одна одної.

- Інтеграція з символьними обчислювальними системами: повна автоматизація символьної верифікації.

- Керована фантазія: дослідження можливості генерації нових гіпотез через контрольоване послаблення фільтрів CE-Protocol.

Теоретична цінність та практична значущість

Теоретична цінність. Дослідження обґрунтовує перехід до «пост-книжкової» ери науки, де теорії стають виконуваними алгоритмами. На відміну від традиційних текстів, що вимагають людської інтерпретації та схильні до деградації (втрата контексту, застарівання нотації), IAI-TE-теорії є самодокументованими та самовідновлюваними системами. Це формує нову епістемологічну парадигму – AI-Resilient Science, де знання стає «безсмертним» завдяки можливості точної регенерації в будь-якому ШІ-середовищі.

Практична значущість. Застосування методу IAI-TE відкриває шлях до:

- Автоматизованої перевірки наукових гіпотез: можливість миттєвої верифікації внутрішньої когерентності нових теорій.

- Прискорення освітнього процесу: генерація персоналізованих навчальних матеріалів з єдиного A-Core.

- Міждисциплінарного синтезу: формалізація зв'язків між теоріями різних галузей через унікальні A-Core.

- Збереження наукової спадщини: конвертація класичних теорій у IAI-TE-формат для захисту від втрати.

Висновки

У ході дослідження було розроблено та апробовано метод індукованого розширення теорії ШІ (IAI-TE), який пропонує принципово новий підхід до побудови наукових знань у симбіозі з великими мовними моделями. Основні результати:

- Подолання логічного дрейфу. Встановлено, що ключовою проблемою використання LLM у фундаментальній науці є прогресуюча втрата когерентності при масштабуванні текстів (частота помилок 3-5/1000 токенів). Запропонована трирівнева архітектура (A-Core, S-Template, Full Specification) у поєднанні з протоколом CE-Protocol редукує логічний дрейф на ~90%, знижуючи частоту помилок до 0.35/1000 токенів.

- Трансформація галюцинацій. Доведено, що за умови використання аксіоматичного ядра як «фільтра реальності», статистичні галюцинації ШІ трансформуються з дефекту на діагностичний інструмент. Систематичні помилки LLM при генерації певних фрагментів сигналізують про недостатню визначеність аксіом або логічні прогалини в самій теорії, перетворюючи ШІ на «логічного опонента».

- Ефективність регенерації знань. На прикладі Темпоральної теорії Всесвіту (TTU) практично підтверджено можливість повної реконструкції складної теоретичної моделі (47 рівнянь, 127 КБ тексту) з мінімального набору даних (~7.2 КБ). Коефіцієнт відновлення формул FRR=1.0, індекс символьної ідентичності SEI=0.94 доводять, що регенерація є дедукцією, а не меморізацією.

- Алгоритмічна епістемологія. Метод IAI-TE закладає підґрунтя для переходу до «пост-книжкової» ери наукових публікацій, де результатом дослідження є не статичний текст, а «виконувана» інтелектуальна система – аналог програмного коду для теоретичного знання. Це формує нову епістемологічну парадигму – AI-Resilient Science, де теорії стають безсмертними завдяки здатності до самовідновлення.

Перспективи подальшого розвитку методу пов'язані з автоматизацією процесу створення аксіоматичних ядер для існуючих наукових парадигм, розробкою систем багатоагентної верифікації знань різними моделями ШІ та дослідженням можливостей «керованої фантазії» – генерації нових наукових гіпотез через контрольоване послаблення фільтрів реальності.

Результати дослідження мають значення не лише для фізики, а й для будь-яких галузей знання з чіткою аксіоматичною структурою: математики, теоретичної хімії, формальної філософії. IAI-TE відкриває новий етап людино-машинного симбіозу, де ШІ виступає не замінником, а підсилювачем людського інтелекту – точним і невтомним «логічним двигуном», керованим творчою інтуїцією дослідника.

Внесок авторів Андрій Лемешко – концептуалізація; методика; Ольга Ткаченко – збір і перевірка емпіричних даних, емпіричне дослідження; Альона Десятко – аналіз джерел, підготовка огляду літератури та теоретичних основ дослідження, висновки.

Декларація про штучний інтелект

Штучний інтелект автори не використовували.

Конфлікт інтересів

Автори заявляють про відсутність конфлікту інтересів та підтверджують, що під час підготовки цієї роботи не існувало жодних комерційних, фінансових чи інших взаємовідносин, які могли б бути розцінені як такі, що здатні вплинути на результати дослідження або їх інтерпретацію. Робота виконана відповідно до принципів академічної доброчесності, етичних норм проведення наукових досліджень та вимог редакційної політики щодо запобігання конфлікту інтересів.

Список використаної літератури

1. Ji Z., Lee N., Frieske R. et al. Survey of hallucination in natural language generation. ACM Computing Surveys. 2023. Vol. 55, No. 12. P. 1–38. DOI: 10.1145/3571730.
2. Lemeshko A. Temporal Theory of the Universe – Minimal Memory Kernel (TTU_CORE_RECALL_v1.0). ResearchGate. 2024. DOI: 10.13140/RG.2.2.28830.40001.
3. Marcus G. The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence. arXiv preprint. 2020. arXiv:2002.06177.
4. Pearl J. Causality: Models, Reasoning, and Inference. 2nd ed. Cambridge: Cambridge University Press, 2009. 484 p.
5. Wolfram S. Writings: On the symbolic and linguistic capabilities of large language models. Wolfram Media. 2023. URL: https://writings.stephenwolfram.com/ (дата звернення: 12.12.2025).
6. Wilkinson M. D. et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. Scientific Data. 2016. Vol. 3, Article 160018. DOI: 10.1038/sdata.2016.18.
7. Krenn M. et al. On Scientific Understanding with Artificial Intelligence. Nature Reviews Physics. 2022. Vol. 4. P. 761–769.
8. Popper K. R. The Logic of Scientific Discovery. London: Routledge, 1959. 544 p.
9. Kuhn T. S. The Structure of Scientific Revolutions. Chicago: University of Chicago Press, 1962. 172 p.
10. Smaragdis E. et al. AI-Driven Knowledge Discovery and Representation in Scientific Domains. AI Magazine. 2023. Vol. 44(4). P. 1–15.
11. Chen M. et al. Evaluating Large Language Models Trained on Code. arXiv preprint. 2021. arXiv:2107.03374.
12. Berners-Lee T., Hendler J., Lassila O. The Semantic Web. Scientific American. 2001. Vol. 284(5). P. 34–43.
13. Stanford Institute for Human-Centered Artificial Intelligence. AI Index Report 2024. Stanford University, 2024. URL: https://aiindex.stanford.edu/report/
14. Vaswani A. et al. Attention Is All You Need. Advances in Neural Information Processing Systems. 2017. Vol. 30. P. 5998–6008.

References

1. Ji Z., Lee N., Frieske R. et al. Survey of hallucination in natural language generation. ACM Computing Surveys. 2023. Vol. 55, No. 12. P. 1–38. DOI: 10.1145/3571730.
2. Lemeshko A. Temporal Theory of the Universe – Minimal Memory Kernel (TTU_CORE_RECALL_v1.0). ResearchGate. 2024. DOI: 10.13140/RG.2.2.28830.40001.
3. Marcus G. The Next Decade in AI: Four Steps Towards Robust Artificial Intelligence. arXiv preprint. 2020. arXiv:2002.06177.
4. Pearl J. Causality: Models, Reasoning, and Inference. 2nd ed. Cambridge: Cambridge University Press, 2009. 484 p.
5. Wolfram S. Writings: On the symbolic and linguistic capabilities of large language models. Wolfram Media. 2023. URL: https://writings.stephenwolfram.com/ (дата звернення: 12.12.2025).
6. Wilkinson M. D. et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. Scientific Data. 2016. Vol. 3, Article 160018. DOI: 10.1038/sdata.2016.18.

7. Krenn M. et al. On Scientific Understanding with Artificial Intelligence. Nature Reviews Physics. 2022. Vol. 4. P. 761–769.

8. Popper K. R. The Logic of Scientific Discovery. London: Routledge, 1959. 544 p.

9. Kuhn T. S. The Structure of Scientific Revolutions. Chicago: University of Chicago Press, 1962. 172 p.

10. Smaragdis E. et al. AI-Driven Knowledge Discovery and Representation in Scientific Domains. AI Magazine. 2023. Vol. 44(4). P. 1–15.

11. Chen M. et al. Evaluating Large Language Models Trained on Code. arXiv preprint. 2021. arXiv:2107.03374.

12. Berners-Lee T., Hendler J., Lassila O. The Semantic Web. Scientific American. 2001. Vol. 284(5). P. 34–43.

13. Stanford Institute for Human-Centered Artificial Intelligence. AI Index Report 2024. Stanford University, 2024. URL: https://aiindex.stanford.edu/report/

14. Vaswani A. et al. Attention Is All You Need. Advances in Neural Information Processing Systems. 2017. Vol. 30. P. 5998–6008.

Надійшла до редакції: 27.11.25
Прийнята до друку: 17.03.26
Опубліковано: 30.03.26