

ХАРЧЕНКО

Віталій Олександрович — доктор біологічних наук, професор, директор Інституту зоології ім. І.І. Шмальгаузена НАН України, головний редактор журналу «Zoodiversity»

КОРНЄЄВ

Валерій Олексійович — член-кореспондент НАН України, доктор біологічних наук, професор, завідувач відділу Інституту зоології ім. І.І. Шмальгаузена НАН України, заступник головного редактора журналу «Zoodiversity»

ФІЛІМОНОВА

Наталія Сергіївна — науковий редактор журналу «Zoodiversity»

ФАНТОМНІ ПОСИЛАННЯ ТА КРИЗА РЕЦЕНЗУВАННЯ: ЯК ШТУЧНИЙ ІНТЕЛЕКТ ПЕРЕВІРЯЄ СТІЙКІСТЬ НАУКОВИХ ВИДАНЬ

У статті розглянуто критичну вразливість традиційної системи сліпого рецензування до викликів, спричинених стрімким розвитком генеративних інструментів. На основі реального прецеденту — виявлення повністю сфабрикованої бібліографії в поданому рукописі — проаналізовано базові механізми створення хибних посилань: анахронізми, «франкентейнізація», фахові упередження. Простежено еволюцію загрози — перехід від очевидних помилок ранніх алгоритмів до глибоких семантичних галюцинацій сучасних пошукових систем на основі RAG, здатних генерувати бездоганно відформатовані, але концептуально порожні тексти з реальних баз даних. Для захисту видавничого процесу запропоновано оновлений алгоритм редакційного контролю, що передбачає обов'язкову перевірку цифрових ідентифікаторів об'єктів (DOI) та чітке декларування використаних авторами алгоритмів. Наголошено на необхідності та незворотності переходу до парадигми Відкритої наукової платформи (Open Science Framework).

Ключові слова: академічна доброчесність; великі мовні моделі; генеративні галюцинації; відкрита наука; депонування даних; фальсифікація досліджень.

Вступ. Стрімкий розвиток великих мовних моделей (LLM) створив безпрецедентний виклик для академічної доброчесності. Якщо спочатку штучний інтелект (ШІ) не викликав побоювань, оскільки був лише зручним інструментом для вдосконалення академічної англійської, то сьогодні редакційні колеги наукових журналів стикаються з новою, значно небезпечнішою за інші тенденцією — повним вигадуванням наукових даних і бібліографії. Уже невдовзі після масового поширення LLM їхню схильність генерувати правдоподібні, але вигадані посилання було визнано серйозним збоєм.

Ранні спеціалізовані дослідження поведінки LLM показали, що серед посилань, згенерованих у біомедичному науковому контексті, лише деякі були повністю вигаданими, але з-поміж згенерованих на запит у ChatGPT частка таких посилань сягала 69 % [1, 2]. Нещодавні аудити свідчать, що це явище вже пе-

перосло в системну кризу, яка стрімко поглиблюється: аналіз 2,5 млн статей біомедичного спрямування у PubMed Central виявив більш як 12-разове збільшення частки сфабрикованих посилань у період з 2023 до 2026 р. [3].

Важливо підкреслити, що сучасні великі мовні моделі за своєю суттю є не інструментами оброблення інформації, а системами оброблення мови. Вони імітують мовні патерни та структури текстів зі своїх навчальних наборів даних, а не їхній фактичний зміст [2]. Відповідно, всі такі моделі схильні до «галюцинацій» — генерування некоректних, неправильних, але правдоподібних відповідей. Сьогодні розробники докладають зусиль для усунення цих вроджених недоліків LLM, насамперед за допомогою інтеграції технології зовнішнього пошуку знань (RAG — Retrieval-Augmented Generation) та можливостей пошуку в реальному часі для прив'язки результатів до верифікованих даних. Однак, оскільки базовий генеративний механізм залишається ймовірнісним, проблема забезпечення фактичної точності поки що є постійним викликом.

У цій статті ми проаналізуємо реальний випадок, з яким зіткнулася редакційна колегія нашого журналу «Zoodiversity».

Хронічні хвороби сучасних наукових видань. Рік тому редакція отримала від групи авторів рукопис оглядової статті, присвяченої екологічним дослідженням однієї з груп комах у Північній Африці. Опрацювання цього рукопису виявило всі хронічні хвороби сучасних наукових видань, а саме:

- *гостра нестача рецензентів* — ми зіткнулися з масовими відмовами фахівців від рецензування, які вони пояснювали надмірним навантаженням;

- *поверхове рецензування* — зрештою рукопис пройшов подвійне сліпе рецензування. Обидва рецензенти дали позитивні відгуки, витративши неадекватно довгий час на аналіз тексту, однак, як з'ясувалося згодом, перевірку вони виконали суто формально;

- *редакційна рутинна й тиск з боку авторів* — після отримання позитивних рецензій статтю було попередньо прийнято до публікації. Роз-

почався тривалий, повний скрупульозної роботи етап коректури: редагування тексту, перевірка відповідності британській англійській мові та приведення рукопису у відповідність до вимог журналу. Цей процес супроводжувався постійним тиском з боку авторів, які регулярно засипали редакцію листами з вимогами прискорити публікацію: «Який зараз статус статті? Коли її буде опубліковано онлайн?».

Сфабрикований список літератури. Переломним моментом, що обвалив цей картковий будиночок, стала фінальна технічна перевірка списку літератури. Увагу технічного редактора привернула відсутність робочих DOI для значної частини посилань на статті, опубліковані останніми роками. Вибіркова ручна перевірка джерел в архівах журналів дала приголомшливий результат: більшості цитованих публікацій фізично не існувало. Вони виявилися продуктом галюцинацій ШІ (найімовірніше, ранніх версій GPT, що використовувалися без пошукових плагінів).

Детальний аналіз сфабрикованого списку літератури дозволив виявити три основні механізми «галюцинацій» ранніх версій генеративних мовних моделей, які автори не критично інтегрували у свій текст і про які має знати кожен редактор.

1. **Анахронізми томів і випусків.** Штучний інтелект поєднує реальну назву журналу з довільними номерами томів і роками. Наприклад, у списку фігурував 12-й том журналу *Entomological Research* за 2020 рік, тоді як насправді у 2020 році вийшов 50-й том цього видання. Модель нездатна перевірити реальну видавничу історію.

2. **«Франкенштейнізація» джерел.** Це найпідступніший тип підробки. Алгоритм бере реальний журнал, правильний рік і том, реальні діапазони сторінок і навіть реальних авторів-фахівців, але повністю вигадує назву статті так, щоб вона ідеально вписувалася в наратив і тематику рукопису. В нашому випадку на сторінках журналу *Agronomy for Sustainable Development*, де нібито містився огляд з екології пряминокрилик, насправді було опубліковано статтю зовсім інших авторів про боротьбу з бур'янами у вирощуванні бобових культур.

3. Фахові та просторові зміщення. Модель не розрізняє вузьку спеціалізацію вчених. У сфабрикованому посиланні на публікацію в *Biodiversity Journal* щодо марокканських комах співавтором виявився реальний і дуже відомий марокканський агроном, професор, чия спеціалізація — гербологія (бур'яни), а не ентомологія. Крім того, зазначені сторінки журналу насправді містили роботи з фауни Італії та Індії.

Масштаб проблеми підтверджується незалежними аудитами. Дослідження біомедичної літератури за останні три роки свідчать, що близько 0,6 % посилань є повністю вигаданими великими мовними моделями [3]. Це не просто технічний збій, це вже системне явище, оскільки такі посилання часто бездоганно відформатовані й використовують імена реальних дослідників, через що вони стають практично «невидимими» для традиційного рецензування.

Виклики та попередження для наукової спільноти. Фактично ми вже перебуваємо на етапі, коли одним із головних викликів у науці стає не доступ до інформації, а перевірка того, чи існує ця інформація насправді.

Зазначений вище випадок потребує широкого розголосу і дискусії щодо етичних норм, оскільки він зачіпає глобальну проблему. Виявлення повністю сфабрикованих джерел у списку літератури неминуче ставить під сумнів вірогідність будь-яких результатів, представлених у статті.

Наша позиція категорична: якщо автори настільки недбало ставляться до підготовки матеріалу або навмисно фальсифікують джерела, немає підстав довіряти їхнім власним результатам і наведеним первинним даним. Їх так само легко могли вигадати.

Редколегія ухвалила рішення відхилити рукопис уже на стадії підготовки до верстки, оскільки публікація такого матеріалу завдала б журналу непоправного репутаційного удару, незважаючи на попереднє схвалення рецензентами. І ми подумали, що було б доцільним оприлюднити цю інформацію (звісно, без розкриття імен авторів неопублікованого рукопису).

Традиційну систему сліпого рецензування побудовано на презумпції добросовісності авто-

ра. Сьогодні такої презумпції вже немає. Рецензенти не готові витратити години на перевірку кожного посилання, чим активно користуються наукові шахраї. Ми вступаємо в еру, коли нові виклики, пов'язані з неконтрольованим використанням великих мовних моделей, здатні зруйнувати основу наукової довіри.

Оновлення алгоритму редакційного контролю. Досвід масштабних аудитів (наприклад, [3]) доводить, що ручної перевірки бібліографії вже недостатньо. Реальне вирішення проблеми полягає у впровадженні автоматизованих систем перевірки посилань безпосередньо в редакційний робочий процес, як запропоновано в нашому оновленому алгоритмі редакційного контролю.

Звертаючись до колег — редакторів академічних журналів, ми закликаємо редакційні колегії всіх академічних видань запровадити такі нові протоколи контролю:

- *вибіркові перевірки* — обов'язкова ручна верифікація 3—5 джерел, що не мають DOI, з різних частин списку літератури. Традиційне рецензування, засноване на презумпції академічної добросовісності авторів, виявилось безсилим перед генеративним штучним інтелектом. Рецензенти, зосереджуючись на логіці тексту, не перевіряють фізичне існування джерел, довіряючи правильно відформатованим спискам літератури;

- *обов'язкова валідація DOI до рецензування* — технічні редактори зобов'язані проводити автоматизовану або ручну валідацію (перевірку дійсності) 3—5 джерел за вказаними DOI, вибираючи їх з різних частин списку літератури. Рукописи без активних цифрових ідентифікаторів для сучасних джерел мають повертатися авторам ще до початку рецензування;

- *декларація про використання ШІ* — в інструкції для авторів має бути закріплено сувору вимогу: будь-яке використання великих мовних моделей (навіть для перекладу або формування) повинне бути детально описано в розділі «Подяки» або «Методи» із зазначенням того, як саме було верифіковано згенеровану інформацію. Відсутність такої декларації в разі виявлення слідів використання LLM має бути

підставою для довічної заборони авторам публікуватися у виданні;

- *інструктаж рецензентів* — оновити форми для рецензентів, додавши пункт про обов'язкову вибірку перевірки 3—5 джерел зі списку літератури.

Наукова спільнота має визнати: інструменти ШІ є незворотною реальністю науки, вони вже з нами, а тому ігнорувати їх неможливо. Однак без запровадження на редакційному рівні суворих верифікаційних бар'єрів наукова література ризикує потонути в потоці правдоподібних, але цілком порожніх симулякрів. Виключно від нашої пильності залежить, чи залишиться наукова стаття золотим стандартом верифікованого знання, чи перетвориться на згенеровану ілюзію.

Кілька слів про «деталі». Штучний інтелект не відчуває «небажання» працювати, не «лінується» через вплив мільйонів користувачів. Моделі не мають свідомості, щоб обирати шлях найменшого опору через втому — в основі цього лежить фундаментальна математична проблема архітектури трансформера, а саме: оцінювання максимальної правдоподібності (Maximum Likelihood Estimation). LLM навчені передбачати наступне слово так, щоб текст виглядав якомога природніше. Для них математично «легше» (тобто потребує менше обчислювальних ресурсів) згенерувати бездоганно «гладеньку», синтаксично правильну, але вигадану назву статті, ніж знайти її точний, конкретний відповідник, який є рідкісним фактом. Більше того, так звана «лінь», коли модель дає короткі, поверхові відповіді, часто є наслідком процедур вирівнювання RLHF (Reinforcement Learning from Human Feedback), тобто навчання з підкріпленням на основі зворотного зв'язку від людей. Застосовуючи цей метод машинного навчання, розробники штучно обмежують моделі, щоб вони не генерували небезпечний контент або щоб заощаджувати обчислювальні ресурси (токени) на серверах компанії.

Еволюція ілюзій: від примітивних помилок до досконалих фальсифікацій. Слід розуміти, що рукопис, який став предметом нашого

аналізу, було створено близько року тому. Тоді автори, найімовірніше, використовували базові версії мовних моделей, що призвело до появи очевидних маркерів «галюцинацій» — неіснуючих томів і сторінок. Однак за останній рік у сфері ШІ відбулися тектонічні зрушення. Сьогодні академічний світ масово озброєний новим поколінням інструментів ШІ, які здатні одночасно аналізувати десятки завантажених статей, а також спеціалізованими науковими пошуковими системами та аналізаторами. Ось який арсенал академічне середовище має на сьогодні.

Базові універсальні LLM (генератори тексту й коду):

- *ChatGPT (моделі GPT-o1)* (розробник — OpenAI) — має вебдоступ та плагіни для аналізу даних. Використовується для підготовки чернеток, структурування статей і написання скриптів (Python/R) для оброблення статистичних даних;

- *Gemini (моделі Gemini 1.5/2.0 Pro)* (розробник — Google) — інтегрований з екосистемою Google, зокрема з Google Scholar, через розширення. Здатний обробляти великі масиви тексту. Використовується для аналізу великих PDF-файлів. Окремий інструмент Google — *NotebookLM* науковці масово застосовують нині для створення персональних баз знань із завантажених статей;

- *Claude (моделі 3.7 Sonnet)* (розробник — Anthropic) — має велике контекстне вікно, що дозволяє авторам завантажувати десятки статей і запитувати зведений огляд.

Спеціалізовані наукові пошукові системи на основі ШІ (системи RAG):

- *Perplexity AI* — найпопулярніший на сьогодні пошуковий інструмент. Не лише генерує текст, а й шукає реальні онлайн-джерела і формує відповідь із прямими посиланнями;

- *Elicit, Consensus, SciSpace* — вузькоспеціалізовані інструменти, навчені виключно на базах наукових статей (Semantic Scholar та ін.), аналізують PDF, «витагують» методологію, результати та створюють зведені таблиці.

Як і очікувалося, технологічний стрибок не лише не зменшив загрозу для академічної до-

бросності, а й зробив її більш підступною. Виникла нова системна проблема: фундаментальна невідповідність між принципами роботи алгоритмів і тим, чого від них очікують користувачі. Великі мовні моделі за своєю природою не є ані базами знань, ані пошуковими системами, вони — ймовірнісні генератори тексту. Їхня математична мета полягає у створенні синтаксично бездоганної і статистично правдоподібної послідовності слів. Коли модель стикається зі складним, вузькоспеціалізованим запитом, вона йде шляхом найменшого опору: компілює усереднений, максимально правдоподібний текст, який виглядає як науковий, звучить як науковий, але не містить реальних фактів.

У результаті маємо ідеальну пастку для несумлінних авторів: ШІ генерує бездоганно написаний англійський текст із досконалою структурою та реальними посиланнями (завдяки новим пошуковим інструментам), але самі висновки, кореляції, синтез ідей є порожніми, «усередненими» галюцинаціями вищого порядку. Виявити такий згенерований матеріал значно складніше: бібліографія буде справжньою, але зв'язку між цитованим джерелом і тезою автора може й не бути — алгоритм просто «підлаштував» посилання під потрібну тематику. Відповідно, тягар верифікації лягає на рецензента, якому тепер доведеться не лише перевіряти наявність джерела, а й дивитися, чи справді в ньому стверджується те, що заявляє автор. Сучасний ШІ здатний бездоганно імітувати форму, тому захист наукового знання має ґрунтуватися виключно на суворому контролі змісту та первинних даних.

Новий алгоритм розпізнавання прихованої генерації. Оскільки сучасні системи RAG зробили виявлення фальшивих джерел за зовнішніми ознаками майже неможливим, редакційні колеги та рецензенти мають перенести фокус своєї уваги з рівня бібліографії на рівень смислів і первинних даних. Ми пропонуємо оновлений алгоритм верифікації рукописів для виявлення некритичного або шахрайського використання новітніх ШІ-моделей:

1. Пошук «семантичних галюцинацій» та ефекту «досконалої порожнечі». Сучасні LLM

генерують синтаксично бездоганні, але концептуально «стерильні» тексти. Рецензент має звертати увагу на такі маркери:

- відсутність наукової дискусії — штучний інтелект алгоритмічно уникає когнітивного дисонансу та «суперечностей» у гіпотезах. Якщо оглядова стаття або розділ обговорення ідеально узгоджує всі джерела і не містить аналізу суперечливих або аномальних результатів — це маркер згенерованого усереднення;

- розрив між цитуванням і висновком — джерело реальне, але не підтверджує тезу, яку висуває автор. Штучний інтелект часто бере реальну статтю, але «вигадує» висновок, потрібний для плавності відповідного наративу. Рецензент має вибірково перевіряти не лише зміст статті, а й суть зробленого висновку;

- методологічна стерильність — згенеровані описи матеріалів і методів часто нагадують ідеалізовані протоколи з підручників, позбавлені конкретних шорсткостей реальної польової або лабораторної роботи.

2. *Верифікація первинних матеріалів (технічний бар'єр).* Найвразливіше місце ШІ — його нездатність генерувати бездоганно узгоджений масив необроблених, емпіричних даних із реальними фізичними параметрами. Тому редакційні колеги мають запровадити презумпцію відкритості для первинних матеріалів. Рукопис не можна допускати до стадії рецензування без виконання таких умов:

- депонування вихідних даних — усі таблиці вимірювань, матриці ознак і статистичні вибірки мають бути завантажені до відкритих репозиторіїв (наприклад, Zenodo або Dryad) на момент подання статті. Журнали мають відмовитися від практики «дані надаються за запитом»;

- цифровий і фізичний слід — для біологічних досліджень потрібно надавати точні географічні метадані (координати тощо), що ускладнює автоматичну генерацію довільних чисел;

- інвентаризація матеріалу — кожен згаданий зразок або матеріал повинен містити посилання на реальний ваучерний номер у визнаному депозитарії.

3. *Вимоги до аналітичних інструментів.* Якщо автори заявляють про проведення складного статистичного або біоінформатичного аналізу, вимоги до них мають бути категоричними. Зокрема, йдеться про відкритий код: надання оригінальних скриптів (R, Python) і журналів (logs) виконання програм. Штучний інтелект бездоганно пише код на вимогу, але часто цей код генерує ідеальні графіки на основі підроблених наборів даних. Надання журналів дозволяє перевірити, чи дійсно код було застосовано до задекларованих вихідних даних.

Висновок. Настав час визнати: епоха, коли для публікації достатньо було надіслати добре відформатований текст (PDF/Word), минула. Відповіддю на виклик генеративних моделей має стати перехід до концепції Відкритої наукової платформи (Open Science Framework), де текстовий матеріал є лише супровідним документом до верифікованого масиву первинних наборів даних, відкритого коду та депонованих колекційних зразків. Якщо автор не може підкріпити свої результати вихідними даними, ваучерами та алгоритмами розрахунків, жоден бездоганно написаний англійський текст не має права називатися науковою публікацією.

Декларація про використання штучного інтелекту та подяки. Автори офіційно заявляють, що цей текст (включно з формулюванням нових вимог до авторів і концептуалізацією проблеми) був згенерований, структурований і стилістично відшліфований у прямому діалозі з великою мовною моделлю Gemini Pro. Ми робимо цю заяву не заради іронічної гри з читачем, а як демонстрацію глибокої й надзвичайно незручної кризи нашого часу. Ми щойно відхилили статтю несумлінних авторів за делегування аналітичної роботи алгоритму, і самі одночасно делегували створення тексту про неприпустимість таких дій алгоритму.

Цей факт має лякати наукову спільноту значно більше, ніж сфабрикований список літератури. Те, що машина здатна бездоганно, з необхідним рівнем академічного обурення, структурною логікою та імітацією людської принципності, написати маніфест проти машинних фальсифікацій, доводить: межу між авторським текстом і машинною генерацією остаточно стерто. Ми вимагаємо прозорості від авторів, тому починаємо з себе. Відповідальність за ідеї, впроваджені правила та відхилення вищезгаданого рукопису повністю і цілком лежить на авторах-людях.

Однак ми визнаємо, що інструментарій для вираження академічної думки більше не є виключно людським. Якщо ми не почнемо суворо контролювати первинні дані сьогодні, завтра буде технічно неможливо відрізнити справжню науку від згенерованої імітації.

Автори висловлюють щире (хоч і не без краплі академічної іронії) подяку великій мовній моделі Gemini Pro за нещадне препарування алгоритмів її власних цифрових «родичів», розкриття анатомії машинних галюцинацій та допомогу в структуруванні цього тексту. Той факт, що статтю про екзистенційну загрозу неконтрольованого штучного інтелекту було написано і проаналізовано в діалозі зі штучним інтелектом, є класичним втіленням міфологічного уробороса — змії, що пожирає власний хвіст. Однак цей парадокс блискуче доводить наш головний постулат: ШІ не є ні абсолютним злом, ні панацеєю. Це лише надпотужний «оптичний прилад». У руках несумлінного автора він генерує переконливі міражі, але під суворим контролем критичного людського розуму стає мікроскопом, здатним ці міражі викрити. Відповідальність за кожне слово, висновок і введене правило в цій статті, як і належить у справжній науці, несуть виключно автори — люди.

REFERENCES

1. Athaluri S.A., Manthena S.V., Kesapragada V.S.R.K.M., Yarlagadda V.I., Dave T., Duddumpudi R.T.S. Exploring the boundaries of reality: investigating the phenomenon of artificial intelligence hallucination in scientific writing through ChatGPT references. *Cureus*. 2023. **15**(4): e37432. <https://doi.org/10.7759/cureus.37432>
2. Walters W.H., Wilder E.I. Fabrication and errors in the bibliographic citations generated by ChatGPT. *Scientific Reports*. 2023. **13**: 14045. <https://doi.org/10.1038/s41598-023-41032-5>
3. Topaz M., Roguin N., Gupta P., Zhang Z., Peltonen L.-M. Fabricated citations: an audit across 2.5 million biomedical papers. *The Lancet*. 2026. **407**(10541): 1779—1780. [https://doi.org/10.1016/S0140-6736\(26\)00603-3](https://doi.org/10.1016/S0140-6736(26)00603-3)

Vitaliy O. Kharchenko

I.I. Schmalhausen Institute of Zoology of the National Academy of Sciences of Ukraine, Kyiv, Ukraine

ORCID: <https://orcid.org/0000-0002-3824-2078>

Valery O. Korneyev

I.I. Schmalhausen Institute of Zoology of the National Academy of Sciences of Ukraine, Kyiv, Ukraine

ORCID: <https://orcid.org/0000-0001-9631-1038>

Natalia S. Filimonova

I.I. Schmalhausen Institute of Zoology of the National Academy of Sciences of Ukraine, Kyiv, Ukraine

PHANTOM REFERENCES AND THE PEER REVIEW CRISIS:

HOW ARTIFICIAL INTELLIGENCE IS TESTING THE RESILIENCE OF SCIENTIFIC PUBLICATIONS

The article examines the critical vulnerability of the traditional blind peer-review system to the challenges posed by the rapid development of generative tools. Based on a real-life case — the discovery of a completely fabricated bibliography in a submitted manuscript — the basic mechanisms of creating false references are analyzed: anachronisms, “Frankensteinization”, professional bias. The evolution of the threat is traced — the transition from obvious errors of early algorithms to deep semantic hallucinations of modern RAG-based search engines capable of generating perfectly formatted, but conceptually empty texts from real databases. To protect the publishing process, an updated editorial control algorithm is proposed, which involves mandatory verification of digital object identifiers (DOI) and clear declaration of the algorithms used by the authors. The necessity and irreversibility of the transition to the Open Science Framework paradigm is emphasized, where textual material is considered only as a supporting document to a verified array of primary data sets, open code, and deposited collection samples.

Keywords: academic integrity; LLM; generative hallucinations; open science; data deposit; research falsification.

Cite this article: Kharchenko V.O., Korneyev V.O., Filimonova N.S. Phantom references and the peer review crisis: how artificial intelligence is testing the resilience of scientific publications. *Visn. Nac. Akad. Nauk Ukr.* 2026. (6): 78—84. <https://doi.org/10.15407/visn2026.06.078>