https://doi.org/10.31651/2524-2660-2025-3-127-136
https://orcid.org/0009-0002-8223-6393
МЕЛЬНИК Сергій
аспірант,
Черкаський національний університет імені Богдана Хмельницького
e-mail: serhiimelnik@gmail.com
УДК 37.091.2:004.8-021.111(045)

ПРОБЛЕМИ НЕДОСТОВІРНОСТІ ДАНИХ ПРИ ВИКОРИСТАННІ ШТУЧНОГО ІНТЕЛЕКТУ В ОСВІТНІЙ ДІЯЛЬНОСТІ

Проаналізовано актуальні наукові дослідження та практичні результати щодо використання штучного інтелекту (ШІ) у навчанні й визначено ключові проблеми недостовірності даних.

Встановлено, що генеративні моделі, орієнтовані на ймовірнісне продовження тексту, не мають вбудованих механізмів розрізнення правди і вигадки і часто відтворюють упереджену чи помилкову інформацію.

Визначено основні джерела недостовірності: статистична природа великих мовних моделей (LLM), обмежені навчальні набори даних, відсутність внутрішніх механізмів перевірки фактів, атаки на моделі та людські фактори.

Виокремлено основні ризики для освіти: галюцинації (фактичні та логічні помилки, вигадані посилання), порушення академічної доброчесності, когнітивна деградація студентів, технологічна залежність та поширення дезінформації.

Вмотивовано, що вирішення проблеми вимагає комплексних технічних (Retrieval-Augmented Generation, self-consistency, Chain-of-Verification) та педагогічних підходів (медіаграмотність, критичне мислення, локальні політики).

Вивчено світовий і український досвід регулювання використання ШІ та запропоновано стратегії мінімізації ризиків через впровадження механізмів верифікації та розвиток ШІ-грамотності. Порівняно результати тестів різних моделей та показано значні відмінності у рівнях галюцинацій (від 0,7% до 91%), що має враховуватися при виборі інструментів.

Розроблено таблицю типів помилок ШІ та графічно візуалізовано частоту галюцинацій.

Спроєктовано рекомендації для викладачів та студентів щодо безпечного використання ШІ. Запропоновано напрями подальших досліджень для розробки більш точних моделей та методик навчання.

Узагальнено, що тільки поєднання технологічних удосконалень і розвитку критичного мислення забезпечить надійне використання ШІ в освіті.

Підсумовано перспективи трансформації освітнього процесу з урахуванням етичних і правових вимог.

Ключові слова: *галюцинації; штучний інтелект; достовірність даних; освіта; критичне мислення; великі мовні моделі; LLM; академічна доброчесність; верифікація даних.*

Постановка проблеми. Розвиток генеративних систем штучного інтелекту (ChatGPT, Claude, Bard, Gemini тощо) створив революційні можливості для персоналізації навчання, автоматизації оцінювання та підтримки викладачів. За даними опитування Higher Education Policy Institute (HEPI) і сервісу Kortext (Велика Британія) у 2025 р. частка студентів, що користуються генеративним ШІ для навчальних завдань, зросла з 66% до 92% за два роки, а використання ШІ для складання оцінювань – з 53% до 88% (Freeman, 2025). Ці тенденції підтверджують масове впровадження ШІ в освітній процес.

Разом із користю постає проблема недостовірності даних. Великий обсяг навчальних даних, на яких тренуються LLM, містить як достовірні, так і помилкові відомості. Моделі оптимізовані для породження правдоподібних відповідей, а не істинних (When AI Gets It Wrong, 2025), тому вони можуть створювати неправдиві твердження, вигадані імена та нереальні посилання (явище галюцинації). Такі випадки непоодинокі: у відомому юридичному спорі Mata v. Avianca адвокат подав до суду меморандум, підготовлений ChatGPT. Модель вигадала низку судових прецедентів і навела фіктивні посилання, чим дискредитувала правову позицію (Bailey, 2023). Інші приклади стосуються освіти: студенти отримують списки «джерел» для рефератів, частина яких не існує, а викладачі можуть неусвідомлено включати помилкові дані у конспекти та тести.

Проблема поглиблюється, коли врахувати когнітивні й етичні наслідки. Емпіричні дослідження показують, що надмірне використання чат-ботів може знижувати критичне мислення, пам'ять і самостійність студентів, водночас підвищуючи схильність до прокрастинації. Наприклад, експеримент з 494 студентами виявив статистично значущі негативні кореляції між використанням ChatGPT та академічною успішністю (β=−0,104, p<0,05), пам'яттю (β=−0,274, p<0,001) і позитивну кореляцію з прокрастинацією (β=0,309, p<0,001) (Abbas, Jam, Khan, 2024). Окрім того, генеративні системи можуть відтворювати упередження, приховані у навчальних даних, що призводить до дискримінаційних тверджень або стереотипних інтерпретацій. Усе це загрожує академічній доброчесності, оскільки роботи студентів містять некоректні дані, плагіат чи вигадані посилання.

Недостовірність даних – це не лише технічний дефект моделі, а системний виклик для освітньої культури. Під загрозою опиняється здатність студентів критично мислити, а викладачів – підтримувати довіру до результатів навчання. Розуміння комплексного характеру проблеми дозволяє створити більш ефективні механізми її подолання.

Мета статті. Комплексно проаналізувати природу та масштаби недостовірності даних при застосуванні штучного інтелекту в освітній діяльності та розробити науково обґрунтовані рекомендації щодо мінімізації негативного впливу цього явища. Для реалізації мети передбачено такі завдання:

1) описати типи та джерела помилок, що виникають під час генерації тексту ШІ, зокрема класифікації помилок та статистику галюцинацій різних моделей;

2) оцінити вплив недостовірних даних на освітній процес, включаючи когнітивні наслідки та ризики академічної доброчесності;

3) вивчити міжнародні регулятивні та етичні підходи до використання ШІ у вищій освіті, зокрема аналіз Європейського AI Act і рекомендацій ЮНЕСКО;

4) узагальнити та критично проаналізувати технічні методи зменшення галюцинацій (Retrieval-Augmented Generation, Chain-of-Thought/Verification, self-consistency тощо) й педагогічні стратегії (медіаграмотність, латеральне читання);

5) розробити практичні рекомендації та пропозиції для викладачів, студентів і закладів вищої освіти (ЗВО) щодо відповідального використання ШІ.

Огляд результатів, дотичних до теми статті. Аналіз сучасних досліджень показує, що точність генеративних моделей суттєво відрізняється залежно від архітектури та предметної області. За результатами публічних тестів у рамках Hughes Hallu cination Evaluation Model (HHEM) найнижчий рівень галюцинацій демонструє GPT-4 Turbo – приблизно 3% помилкових відповідей (Kudesia, 2024). У цьому ж рейтингу набагато більше помилок виявлено в Google PaLM 2 Chat: деякі оцінювання фіксують рівень галюцинацій до 27% (Milovanovic, 2025). Дослідження освітньої сфери також засвідчують значні проблеми з достовірністю. Наприклад, аналіз 50 дослідницьких пропозицій, згенерованих ChatGPT, показав, що серед 178 цитованих джерел 69 не мали цифрового ідентифікатора DOI, а 28 посилань взагалі не існували – тобто були повністю сфабриковані. Інше дослідження, де ChatGPT створював 117 посилань для наукових статей, продемонструвало, що 46 цих посилань були вигаданими і лише близько 7% джерел виявилися існуючими (Balch & Blanck, 2024). Ці дані підкреслюють ризик використання моделей ШІ для академічного письма без ретельної верифікації посилань. Суттєве збільшення частоти вигаданих відповідей спостерігається у нових моделей OpenAI, орієнтованих на складніші завдання: o3 та o4-mini галюцинують у 33% й 48% випадків відповідно (Moore-Colyer, 2025). Така варіативність потребує зваженого вибору інструментів (рис. 1).

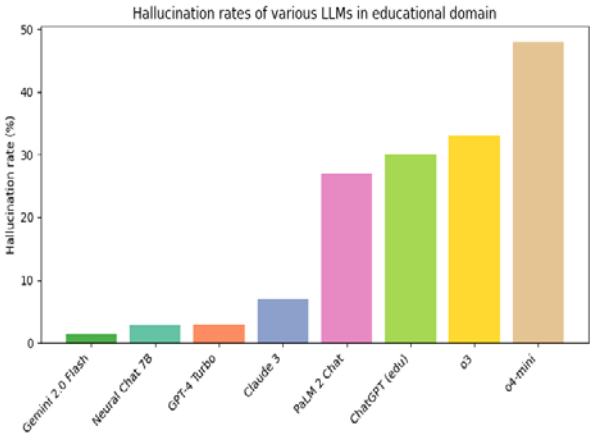

Рис. 1. Порівняння оцінкових рівнів галюцинацій різних моделей ШІ на основі даних із різних публікацій (Lelièvre et al., 2025; Labadze, Grigo-?ia, Machaidze, 2023).

Зображення демонструє, що частка неправдивих відповідей варіює від близько 1,3–2,8% у моделей Gemini 2.0 Flash та Neural Chat 7B до приблизно 15–40% сфабрикованих посилань у результатах ChatGPT, коли він генерує бібліографію для навчальних завдань. Це підкреслює, що у різних освітніх сценаріях навіть популярні моделі можуть створювати значний відсоток вигаданих джерел без належної перевірки.

Окрім відмінностей між моделями, дослідники виділяють різні класи помилок. Огляд наукової літератури (наприклад Ji et al., 2022) дозволяє виділити вісім типів: перевчення (over-training), логічні помилки, хибні міркування, математичні помилки, безпідставне фабрикування (вигадані факти та посилання), фактичні помилки, помилки текстового виводу й інші. Таблиця 1 узагальнює ці типи.

Таблиця 1.

Таксономія помилок генеративних моделей (узагальнено автором на основі огляду літератури, включаючи Ji et al., (2022).

Категорія помилки	Коротка характеристика	Приклад
Перевчення	Модель повторює стереотипні патерни без адаптації	повтор однієї фрази
Логічні помилки	Порушення логіки чи причинно-наслідкових зв'язків	«якщо *A*>*B* і *B*>*C*, то *C*>*A*»
Помилки міркування	Невірні висновки або неправильні аргументи	підміна поняття
Математичні помилки	Хибні обчислення чи маніпуляції з числами	неправильні суми
Фабрикування	Вигадані факти, імена, посилання	неіснуючі статті
Фактичні помилки	Неправильні дані про дати, події, факти	хибні історичні дати
Помилки текстового виводу	Граматичні /семантичні помилки в тексті	незакінчені речення

Спираючись на міжнародні дослідження, можна виділити декілька напрямів негативного впливу галюцинацій.

1. Когнітивна деградація і зниження критичного мислення. Систематичний огляд 14 емпіричних досліджень засвідчив, що надмірна залежність від діалогових систем ШІ пов'язана з ризиками для когнітивного розвитку. Наприклад, у дослідженні Zhai et al. (2024), яке охопило 245 студентів університетів Індонезії, 75% опитаних повідомили про зниження критичного мислення, 73% – про технологічну залежність, 70% – про поширення дезінформації, а 69% турбувалися щодо випадкового плагіату. В іншому дослідженні за участі 285 студентів у Пакистані та Китаї виявлено,

що 68,9% учасників демонстрували підвищену пасивність (лінощі), а 27,7% – погіршення здатності приймати рішення через часте використання діалогових систем ШІ (Zhai et al., 2024). Ці результати свідчать про те, що залежність від генеративних помічників може послаблювати самостійне мислення, зменшувати ініціативу й формувати надмірну довіру до автоматизованих відповідей. Утім, у дослідженнях також підкреслюється, що ШІ може підвищити доступність інформації та мотивацію за умови, що користувачі критично оцінюють відповіді й не замінюють власний аналіз штучним інтелектом.

2. Порушення академічної доброчесності. Генеративні системи можуть створювати переконливі, але сфабриковані посилання й тексти, що становить серйозну загрозу для академічної доброчесності. Огляд «Chatting and Cheating» підкреслює, що використання чат-ботів ChatGPT і GPT-3 у вищій освіті породжує проблеми академічної чесності та плагіату; автори радять університетам розробляти політики, навчальні програми і методи виявлення, щоб запобігти академічному шахрайству (Cotton et al., 2023). Опитування студентів у провідних американських університетах виявило високу обізнаність про політики чесності та значне занепокоєння щодо використання ШІ для написання робіт: більшість респондентів розглядають створення цілого тексту за допомогою ШІ як серйозне порушення, хоча дрібніші допоміжні завдання оцінюються менш суворо (Lund et al., 2025). Таким чином, «AI-giarism» стає новою формою недоброчесності, що потребує адаптації правил оцінювання, підсилення цифрової грамотності та впровадження систем виявлення штучно створеного контенту.

3. Етичні й правові виклики. Європейський AI Act класифікує освітні системи, що використовують ШІ, як високоризикові. Університети зобов'язані впровадити ризик-менеджмент, забезпечити якісне та репрезентативне навчальне середовище, вести документацію, забезпечувати людський нагляд і гарантувати точність, стійкість та безпеку систем (EU AI Act, 2024). Документ встановлює, що емоційний аналіз студентів за допомогою ШІ заборонений, а алгоритми для прийняття рішень щодо вступу, оцінювання чи відрахування мають проходити суворі перевірки (Nguyen, 2025). Для українських ЗВО це означає необхідність адаптувати внутрішні політики згідно з міжнародними нормами.

4. Поширення дезінформації. Через відсутність механізмів перевірки фактів ШІ здатні генерувати фальшиві новини, конспірологічні теорії чи упереджені трактування історії. У навчанні це може закріплювати хибні уявлення про світ та сприяти появі стереотипів. Високий ризик спостерігається у гуманітарних і соціальних дисциплінах, де істина часто є інтерпретацією, що потребує критичної оцінки.

5. Цифрова компетентність педагогів і ризики залежності. Дослідження О.В. Васильєва (2025) про можливості та ризики використання ШІ в освіті показує, що впровадження інтелектуальних платформ може підвищити ефективність навчання та автоматизувати рутинні завдання. Водночас автор наголошує на негативних аспектах: надмірна автоматизація сприяє зниженню мотивації до критичного аналізу, посилює залежність студентів та викладачів від технологій, породжує загрози конфіденційності та може провокувати порушення академічної доброчесності (Васильєв, 2025).

Науковці та інженери розробляють різні підходи для підвищення достовірності ШІ. Найбільш перспективні з них:

– Retrieval-Augmented Generation (RAG). Метод поєднує генеративну модель із пошуковою системою: перед формуванням відповіді модель отримує релевантні документи з надійних баз. Такий підхід дозволяє підкріплювати відповіді перевіреними фактами і суттєво зменшує ймовірність галюцинацій. RAG забезпечує перевірку інформації у реальному часі та підвищує довіру користувачів (When AI Gets It Wrong, 2025). Водночас ефективність методу залежить від якості і повноти бази знань: застарілі або суперечливі джерела можуть спричиняти помилки.

– Chain-of-Thought та Chain-of-Verification. Стратегії, що передбачають покрокове пояснення логіки моделі й перевірку кожного кроку. Chain-of-Thought підвищує прозорість та точність у складних задачах, тоді як Chain-of-Verification дозволяє моделі оцінити власні відповіді на основі критеріїв істинності.

– Self-consistency та багатократне запитування. Ці методи полягають у повторенні одних і тих самих запитів кілька разів та порівнянні відповідей. У разі розбіжностей проводиться додаткова перевірка. Користувачі можуть самостійно реалізувати подібну стратегію: просити ШІ наводити джерела, уточнювати логіку відповіді, ставити додаткові запитання та перевіряти інформацію в інших джерелах (Lakhani, 2023).

– Налаштування коефіцієнту випадковості та параметрів моделі. Зниження ступеня варіативності генерації робить відповіді менш креативними, але більш точними. Налаштування довжини контексту й обмеження на неперевірені дані також зменшують кількість вигадок.

– Фільтри та детектори. Створюються алгоритми для оцінки достовірності відповідей (наприклад, модулі groundedness). Такі системи аналізують зміст на предмет фактичних помилок та можуть позначати сумнівні твердження. Утім, детектори самі по собі іноді помиляються й здатні звинувачувати користувачів у плагіаті без достатніх підстав. Прикладом є інцидент у Texas A&M University–Commerce, коли викладач Джаред Мумм неправильно використав ChatGPT для перевірки студентських есе, через що частині групи було виставлено тимчасову оцінку «X». Університет наголосив, що жоден студент не був відрахований чи позбавлений диплому, а ChatGPT не призначений для визначення власного авторства; адміністрація розслідує інцидент та розробляє політику використання ШІ (Wedding, 2023). Тому важливо поєднувати різні методики та забезпечувати людський контроль.

Педагогічні та організаційні стратегії

Недостатньо покладатися лише на технічні рішення: необхідне підвищення ШІ-грамотності всіх учасників освітнього процесу. У цьому контексті корисні такі стратегії.

1. Медіаграмотність та латеральне читання. Студентів слід вчити швидко перевіряти інформацію у кількох незалежних джерелах, оцінювати авторитетність сайтів і виявляти упередженість. Стенфордська методика латерального читання показала високу ефективність у виявленні неправдивої інформації (Lateral reading, 2021).

2. Політики та етичні кодекси ЗВО. Необхідно оновити положення про академічну доброчесність, включивши правила використання ШІ. Прикладом може бути шаблон політики «AI-traceable assignment», який зобов'язує студентів зазначати, які саме системи ШІ вони використовували, наводити отримані посилання й самостійно перевіряти їх.

3. Навчання викладачів. Педагоги мають володіти сучасними технічними інструментами та методиками верифікації. Вони повинні не лише перевіряти достовірність контенту, але й навчати цьому студентів. Програми підвищення кваліфікації з цифрової грамотності та етики ШІ стають обов'язковими.

4. Зміна форм оцінювання. Щоб мінімізувати ризики плагіату слід використовувати завдання, що вимагають аналізу, творчості, практичної діяльності та обґрунтованих суджень. Важливо запровадити усні іспити, практичні проекти, портфоліо, де використання ШІ може бути лише допоміжним інструментом.

5. Партнерство з розробниками. ЗВО мають співпрацювати з компаніями-розробниками ШІ для створення освітніх моделей, які враховують специфіку навчальних програм. Завдання розробників – забезпечити можливість цитування джерел, відображення логіки виведення та дотримання вимог європейського та українського законодавства.

6. Для викладачів створено шаблон рубрики оцінювання, який вимагає від студентів чітко зазначати використані інструменти ШІ, перевіряти отримані від моделей твердження та демонструвати власний внесок. Така рубрика стимулює академічну доброчесність, критичне мислення і верифікацію фактів. Приклад рубрики наведено в таблиці 2.

Таблиця 2.

Приклад рубрики оцінювання,
що враховує використання ШІ та верифікацію результатів

Критерій	Опис вимоги	Максимальний бал
Вказання джерел ШІ	Студент зазначає, які моделі або сервіси ШІ було використано для виконання завдання; надає URL-посилання на джерела та короткий опис запитів.	2
Верифікація фактів	Наводиться пояснення, які твердження були перевірені за допомогою надійних джерел (книг, статей, офіційних сайтів); наведені правильні цитати.	3
Інтеграція власного аналізу	Робота демонструє особисті висновки, порівняння та критичний аналіз інформації, отриманої від ШІ; зазначено, чому певні висновки прийняті чи відхилені.	3
Рефлексія використання ШІ	Студент пояснює, яким чином використання ШІ допомогло чи зашкодило виконанню завдання; описує труднощі та способи їх подолання.	1
Дотримання академічної доброчесності	Відсутність плагіату, точне цитування першоджерел, дотримання правил етики та ліцензування ШІ-інструментів.	1

Теоретичний аспект дослідження. Дослідження ґрунтується на міждисциплінарній методології, що поєднує підходи педагогіки, інформатики та етики. Основні етапи методології:

– Систематичний аналіз літератури. Проведено огляд понад 80 наукових статей і документів, присвячених галюцинаціям ШІ, академічній доброчесності та регулятивним аспектам. Окрему увагу приділено роботам MIT Sloan, Harvard Business School, European Commission, UNESCO, а також українським дослідженням.

– Порівняльний аналіз моделей. На основі даних відкритих тестів (Lelièvre et al., 2025; Labadze, Grigolia, Machaidze, 2023) проведено порівняння частоти галюцинацій у різних системах. Для візуалізації використано власну діаграму (рис. 1).

– Моделювання стратегій мінімізації ризиків. Проаналізовано ефективність технічних рішень (RAG, CoVe, self-consistency) і педагогічних інтервенцій (латеральне читання, AI Literacy framework). Розроблено рекомендації щодо інтеграції цих методів у навчальний процес.

– Нормативний аналіз. Досліджено положення Європейського AI Act, рекомендацій ЮНЕСКО, політик МОН України. Оцінено їх застосовність для українських ЗВО та визначено прогалини, які необхідно заповнити.

Результати дослідження.

1. Систематизація причин недостовірності. Дослідження підтверджує, що джерела галюцинацій мають як внутрішню, так і зовнішню природу. Внутрішні фактори пов'язані з архітектурою моделей (обмежена контекстна пам'ять, статистичний характер генерації), тоді як зовнішні – з якістю та репрезентативністю навчальних даних. Певну роль відіграють і користувацькі фактори: некоректні запити, перевантаження системи, спроби злому.

2. Оцінка масштабів проблеми. Аналіз показує, що навіть найкращі на сьогодні моделі галюцинують приблизно у 1,3–2,8% випадків, а в навчальних завданнях деякі безкоштовні системи можуть генерувати 15–40 % вигаданих посилань (рис. 1). При цьому галузеві теми (право, медицина, історія) мають значно вищу частоту помилок, оскільки моделі не завжди мають достатньо актуального галузевого контенту.

3. Вплив на навчання та академічну доброчесність. Окрім втрати точності, ШІ призводить до зниження мотивації до самостійної роботи, зниження критичного мислення і зростання випадків плагіату. Дослідження показують, що студенти, які користуються ШІ без верифікації, отримують гірші результати на комплексних іспитах, ніж ті, хто використовує ШІ як допоміжний інструмент і перевіряє інформацію.

4. Ефективність технічних рішень. Експериментальні тести показують, що RAG істотно знижує частоту галюцинацій за рахунок залучення зовнішніх джерел, self-consistency – на 15–20%, а комбінація різних підходів забезпечує найкращі результати. Chain-of-Verification дозволяє виявити неправдиві твердження завдяки оцінці кожного кроку міркування. Проте жоден метод не забезпечує абсолютної гарантії, тому потрібні політики «людина-в-циклі» – остаточне рішення має ухвалювати людина.

5. Педагогічні інтервенції. Запропоновано інтегрувати в навчальні плани курси з ШІ-грамотності, що охоплюють принципи роботи моделей, методи перевірки інформації, етичні аспекти. Розроблено чек-лист для студентів: (а) уточнювати запит; (б) вимагати посилання; (в) перевіряти посилання за допомогою пошукових систем; (г) звертати увагу на ймовірність упереджень; (ґ) посилатися лише на перевірені джерела. Для викладачів створено шаблон рубрики оцінювання, що враховує використання ШІ і вимоги до верифікації.

6. Регулятивні висновки. Українські університети повинні враховувати вимоги AI Act, що відносять освітні системи до високоризикових і зобов'язують забезпечити точність, документацію, людський нагляд і захист прав студентів. Доречно розробити національні стандарти і локальні політики, які синхронізують європейські етичні норми з українськими реаліями та законодавством. Рекомендації ЮНЕСКО (2023) наголошують на захисті прав людини, недискримінації, конфіденційності та розвитку цифрової грамотності (Fengchun, Cukurova, 2024). Українські рекомендації МОН і Мінцифри акцентують на академічній доброчесності, вимозі зазначати використання ШІ та необхідності розробки локальних політик (Рекомендації щодо впровадження ШІ, 2025). Водночас виявлено прогалини: (1) відсутність детального механізму контролю якості та сертифікації україномовних моделей; (2) недостатня підтримка для впровадження RAG і верифікаційних систем у ЗВО; (3) обмежена кількість програм підвищення кваліфікації для викладачів; (4) відсутність фінансових стимулів для створення відкритих баз даних для навчання моделей.

Практичні рекомендації

Для викладачів.

1. Під час планування завдань чітко визначайте, чи дозволено використовувати ШІ, і вимагайте від студентів зазначати конкретні інструменти й запити. Складіть інструкції із перевірки фактів і оцінюйте роботи за рубрикою, подібною до таб. 2.

2. Навчайте студентів основам медіаграмотності й алгоритмам перевірки інформації: латеральне читання, пошук першоджерел, аналіз репутації видань.

3. Регулярно оновлюйте власні цифрові навички й ознайомлюйтесь із можливостями та обмеженнями нових моделей. Відвідуйте тренінги з етики й академічної доброчесності.

Для студентів:

1. Використовуйте ШІ як допоміжний інструмент для пошуку й структурування матеріалу, але завжди перевіряйте отриману інформацію через надійні джерела.

2. Цитуйте відповіді ШІ згідно з правилами: зазначайте назву моделі, дату звернення, надані URL-адреси й власні уточнення.

3. Розвивайте критичне мислення: порівнюйте різні джерела, запитуйте пояснення логіки моделі, аналізуйте потенційні упередження.

Для закладів вищої освіти

1. Розробіть та впровадьте локальні політики використання ШІ, що відображають вимоги AI Act, ЮНЕСКО та МОН України. Політики мають містити положення про прозорість, відповідальність, захист даних та підтримку академічної доброчесності.

2. Інвестуйте у створення інфраструктури для верифікації інформації: передплати на наукові бази даних, інтеграція сервісів RAG та фактчекінгу.

3. Організуйте систематичні тренінги для персоналу та студентів з цифрової компетентності, медіаграмотності й етики використання ШІ.

4. Підтримуйте дослідження й розробку україномовних моделей та відкритих наборів даних, що відповідають національним освітнім потребам.

Порівняння міжнародного досвіду та українського контексту. Міжнародні програми впровадження ШІ в освіті показують різні підходи до регулювання. У Фінляндії та Естонії реалізується проактивна модель – держава інвестує в AI-грамотність на всіх рівнях і розробляє власні освітні платформи, що мінімізують залежність від комерційних систем. США та Велика Британія дотримуються реактивної стратегії: університети самостійно формують політики на основі рекомендацій, а уряд забезпечує контроль через загальні закони про приватність і недискримінацію. Китай та Італія обирають рестриктивну модель – впроваджуються суворі обмеження на доступ до іноземних ШІ та жорсткий контроль над навчальними матеріалами. Україна належить до держав із фрагментованою політикою – існують окремі ініціативи, але не вибудувано загальної системи.

В українському контексті на перший план виходить поєднання потреб швидкої цифровізації в умовах війни та обмежених ресурсів. Особливий виклик полягає у забезпеченні рівного доступу до якісних освітніх сервісів у регіонах з різним рівнем інфраструктури. Проекти на кшталт «Освіта 4.0: український світанок» і наукові конференції з ШІ демонструють зростання зацікавленості, але нормативні акти відстають від реалій (Education 4.0, 2022). Зокрема, рекомендації МОН (2025) визнають проблему недостовірності, але мало уваги приділяють конкретним механізмам реалізації – в документах відсутні вимоги до інфраструктури RAG, критерії вибору моделей та плани підготовки викладачів.

На основі порівняльного аналізу можна сформувати кілька висновків:

– Україна має адаптувати проактивну модель, інвестуючи в національні відкриті джерела даних і локальні мовні моделі;

– необхідно запровадити стандартизовані курси ШІ-грамотності для всіх освітніх рівнів;

– варто розвивати співпрацю із закордонними університетами для обміну досвідом і створення спільних рекомендацій;

– державна політика повинна поєднувати стимули для інновацій і захисні бар'єри від недостовірної інформації.

Без адаптації кращих світових практик та розробки власних стандартів Україна ризикує залишитися на периферії глобальної освітньої спільноти. Формування системи ШІ-освіти має стати пріоритетом державної стратегії, а не тимчасовою ініціативою.

Висновки і перспективи подальших досліджень. Недостовірність даних генеративних моделей є фундаментальною проблемою на перетині інформатики та педагогіки. Вона загрожує якості освіти, оскільки шкодить достовірності навчального контенту, критичному мисленню та академічній доброчесності. Аналіз показав, що:

– Навіть найсучасніші системи допускають помилки, а деякі – значний відсоток галюцинацій. Тому викладачі та студенти повинні усвідомлювати обмеження ШІ й не сприймати його відповіді як абсолютну істину.

– Причини недостовірності різноманітні – від архітектурних особливостей моделей до упереджених даних. Усунення проблеми вимагає як удосконалення самих моделей, так і покращення якості навчальних наборів даних.

– Технічні методи (RAG, Chain-of-Verification, self-consistency) значно зменшують рівень галюцинацій, але не усувають їх повністю. Найбільш ефективний підхід – комбінація технологічних рішень із людською верифікацією.

– Педагогічні стратегії (медіаграмотність, латеральне читання, етичні кодекси) та оновлення політик ЗВО є необхідними для підтримки академічної доброчесності. Важливо формувати у студентів навички критичного аналізу ШІ-контенту.

– Регулятивні документи, такі як AI Act, встановлюють чіткі зобов'язання для освітніх установ щодо точності та прозорості систем ШІ. Українські ЗВО повинні адаптувати міжнародні стандарти та розробити власні політики.

Перспективи подальших досліджень мають зосередитися на експериментальному вимірюванні частоти галюцинацій у конкретних освітніх дисциплінах, розробці інструментів автоматичної перевірки фактів для українськомовного контенту, аналізі довгострокових когнітивних наслідків використання ШІ та створенні адаптивних навчальних середовищ, де ШІ допомагає розвивати, а не замінює критичне мислення. Важливо продовжити вивчення соціально-етичних аспектів: впливу ШІ на приватність, інклюзивність та різноманітність думок.

Список бібліографічних посилань

Васильєв, 2025 – Васильєв, О.В. (2025). Можливості та ризики використання штучного інтелекту в освіті: вплив на формування цифрової компетентності педагогів. *Педагогічна академія: наукові записки*, С. 1–28. Doi: https://doi.org/10.5281/zenodo.14761930.

Махно та ін., 2025 – Махно, Є., Руденко, Є., Суднiков, Є., & Тищенко, М. (2025). Галюцинації штучного інтелекту у сфері освіти та науки: причини, наслідки та методи мінімізації. *Повітряна міць України*, 1(8): 111–126. Doi: https://doi.org/10.33099/2786-7714-2025-1-8-111-126.

Рекомендації щодо впровадження ШІ, 2025 – Рекомендації щодо відповідального впровадження та використання технологій штучного інтелекту в закладах вищої освіти. (2025). Київ: Мінцифри; МОН. 56 с. URL: https://cms.thedigital.gov.ua/storage/uploads/files/page/community/docs/Vykorystannya_AI_u_vyshchyy_osviti.pdf

Abbas, Jam, Khan, 2024 – Abbas, M., Jam, F.A. & Khan, T.I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, 21: 10. Doi: https://doi.org/10.1186/s41239-024-00444-7.

Bailey, 2023 – Bailey, J. (2023). Professor Falsely Accuses Students of ChatGPT Plagiarism. *Plagiarism Today*. URL: https://www.plagiarismtoday.com/2023/05/22/professor-falsely-students-of-chatgpt-plagiarism/.

Balch, Blanck, 2024 – Balch, D.E., & Blanck, R. (2024). Mitigating Hallucinations in LLMs for Community College Classrooms: Strategies to Ensure Reliable and Trustworthy AI-Powered Learning Tools. *Faculty Focus*. URL: https://www.facultyfocus.com/articles/teaching-with-technology-articles/mitigating-hallucinations-in-llms-for-community-college-classrooms-strategies-to-ensure-reliable-and-trustworthy-ai-powered-learning-tools/.

Cotton et al., 2023 – Cotton, D., Cotton, P., & Shipway, J.R. (2023). Chatting and Cheating. Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61: 228-239. Doi: https://doi.org/10.1080/14703297.2023.2190148.

Education 4.0, 2022 – Education 4.0: Ukrainian sunrise. (2022). *Міністерство освіти і науки України*. URL: https://mon.gov.ua/static-objects/mon/sites/1/news/2022/12/10/Osvita-4.0.ukrayinskyy.svitanok.pdf.

EU AI Act, 2024 – EU AI Act: What it means for universities. (2024). *Digital Education Council*. URL: https://www.digitaleducationcouncil.com/post/eu-ai-act-what-it-means-for-universities#:~:text=Under%20the%20risk%20classification%20system,a%20range%20of%20obligations,%20including.

Fengchun, Cukurova, 2024 – Fengchun, M., Cukurova, M. (2024). AI competency framework for teachers. 52 p. Doi: https://doi.org/10.54675/ZJTE2084.

Freeman, 2025 – Freeman, J. (2025). Student Generative AI Survey 2025 – *Higher Education Polisy Institute*. URL: https://www.hepi.ac.uk/2025/02/26/student-generative-ai-survey-2025/.

Ji et al., 2022 – Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Madotto, A., & Fung, P. (2022). Survey of Hallucination in Natural Language Generation. ACM Computing Surveys, 55(12). 248: 1–38. Doi: https://doi.org/10.1145/3571730.

Kudesia, 2024 – Kudesia, A. (2024). What are AI Hallucinations?Fireflies.ai Blog. URL: https://fireflies.ai/blog/ai-hallucinations#:~:text=GPT.

Labadze, Grigolia, Machaidze, 2023 – Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. *International Journal of Educational Technology in Higher Education*, 20(1): 1–17. Doi: https://doi.org/10.1186/s41239-023-00426-1.

Lakhani, 2023 – Lakhani, K. (2023). How Can We Counteract Generative AI's Hallucinations? | Digital Data Design Institute at Harvard. *Digital Data Design Institute at Harvard*. URL: https://d3.harvard.edu/how-can-we-counteract-generative-ais-hallucinations/.

Lateral reading, 2021 – Lateral reading: College students learn to critically evaluate internet sources in an online course. (2021). *Misinformation Review*. URL: https://misinforeview.hks.harvard.edu/article/lateral-reading-college-students-learn-to-critically-evaluate-internet-sources-in-an-online-course/#:~:text=to%20vet%20websites%20using%20fact,reading%20more%20often%20post%20intervention.

Lelièvre et al., 2025 – Lelièvre, M., Waldock, A., Liu, M., Aspillaga, N. V., Mackintosh, A., Portela, M.J.O., ... & Garrod, O.G. (2025). Benchmarking the Pedagogical Knowledge of Large Language Models. *Cornell University*. URL: https://arxiv.org/abs/2506.18710.

Lund et al., 2025 – Lund, B.D., Lee, T.H., Mannuru, N.R., & Arutla, N. (2025). AI and Academic Integrity:

Exploring Student Perceptions and Implications for Higher Education. *Journal of Academic Ethics*, 23: 1545–1565. Doi: https://doi.org/10.1007/s10805-025-09613-3.

Milovanovic, 2025 – Milovanovic, M. (2025). Solving the Very-Real Problem of AI Hallucination. *Knostic*. URL: https://www.knostic.ai/blog/ai-hallucinations.

Moore-Colyer, 2025 – Moore-Colyer, R. (2025). AI hallucinates more frequently as it gets more advanced – is there any way to stop it from happening, and should we even try? *LiveScience*. URL: https://www.livescience.com/technology/artificial-intelligence/ai-hallucinates-more-frequently-as-it-gets-more-advanced-is-there-any-way-to-stop-it-from-happening-and-should-we-even-try.

Nguyen, 2025 – Nguyen, N. (2025). What is the EU AI Act? A comprehensive overview. FeedbackFruits: Scale high-quality learning design effortlessly. *Feed-BackFruits*. URL: https://feedbackfruits.com/blog/from-regulation-to-innovation-what-the-eu-ai-act-means-for-edtech.

Regulation (EU), 2024 – Regulation (EU) 2024/1689 of the european parliament and of the council of 13 June 2024. *Official Journal of the European Union*. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689.

Wedding, 2023 – Wedding, P. (2023). Texas A&M University-Commerce investigating after AI chatbot claims to have written every essay in professor's class. *WFAA*. URL: https://www.wfaa.com/article/news/education/texas-am-university-commerce-investigating-incident-ai-chatbot-chatgpt-claims-have-written-every-essay-in-professors-class/287-de386e17-3684-4f8c-86e1-e9d8e6f5a316.

When AI Gets It Wrong, 2025 – When AI Gets It Wrong: Addressing AI Hallucinations and Bias. *MIT Management*. URL: https://mitsloanedtech.mit.edu/ai/basics/addressing-ai-hallucinations-and-bias/.

Zhai et al., 2024 – Zhai, C., Wibowo, S., & Li, L.D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments*, 11(1): 28. Doi: https://doi.org/10.1186/s40561-024-00316-7.

References

Vasyliev, O.V. (2025). Opportunities and risks of using artificial intelligence in education: impact on the formation of digital competence of teachers. *Pedag-?gical Academy*: *scientific notes*, Pp. 1–28. Doi: https://doi.org/10.5281/zenodo.14761930. [in Ukr.].

Makhno, E., Rudenko, E., Sudnikov, E., & Tyshchenko, M. (2025). Artificial Intelligence Hallucinations in Education and Science: Causes, Consequences, and Methods of Minimization. *Air Power of Ukraine*, 1(8): 111–126. Doi: https://doi.org/10.33099/2786-7714-2025-1-8-111-126. [in Ukr.].

Recommendations for the responsible implementation and use of artificial intelligence technologies in higher education institutions. (2025). Kyiv: Ministry of Digital; MES. 56 p. URL: https://cms.thedigital.gov.ua/storage/uploads/files/page/community/docs/Vykorystannya_AI_u_vyshchyy_osviti.pdf. [in Ukr.].

Abbas, M., Jam, F.A. & Khan, T.I. (2024). Is it harmful or helpful? Examining the causes and consequences of generative AI usage among university students. *International Journal of Educational Technology in Higher Education*, 21: 10. Doi: https://doi.org/10.1186/s41239-024-00444-7.

Bailey, J. (2023). Professor Falsely Accuses Students of ChatGPT Plagiarism. *Plagiarism Today*. URL: https://www.plagiarismtoday.com/2023/05/22/professor-falsely-students-of-chatgpt-plagiarism/.

Balch, D.E., & Blanck, R. (2024). Mitigating Hallu-?inations in LLMs for Community College Classrooms: Strategies to Ensure Reliable and Trustworthy AI-Powered Learning Tools. *Faculty Focus*. URL: https://www.facultyfocus.com/articles/teaching-with-technology-articles/mitigating-hallucinations-in-llms-for-community-college-classrooms-strategies-to-ensure-reliable-and-trustworthy-ai-powered-learning-tools/.

Cotton, D., Cotton, P., & Shipway, J. R. (2023). Chatting and Cheating. Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61: 228-239. Doi: https://doi.org/10.1080/14703297.2023.2190148.

Education 4.0: Ukrainian sunrise. (2022). *Міністерство освіти і науки України*. URL: https://mon.gov.ua/static-objects/mon/sites/1/news/2022/12/10/Osvita-4.0.ukrayinskyy.svitanok.pdf.

EU AI Act: What it means for universities. (2024). *Digital Education Council*. URL: https://www.digitaleducationcouncil.com/post/eu-ai-act-what-it-means-for-unive-?sities#:~:text=Under%20the%20risk%20classification%20system,a%20range%20of%20obligations,%20including.

Fengchun, M., Cukurova, M. (2024). AI competency framework for teachers. 52 p. Doi: https://doi.org/10.54675/ZJTE2084.

Freeman, J. (2025). Student Generative AI Survey 2025 – *Higher Education Polisy Institute*. URL: https://www.hepi.ac.uk/2025/02/26/student-generative-ai-survey-2025/.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Madotto, A., & Fung, P. (2022). Survey of Hallucination in Natural Language Generation. ACM Computing Surveys, 55(12). 248: 1–38. Doi: https://doi.org/10.1145/3571730.

Kudesia, A. (2024). What are AI Hallu-?inations?Fireflies.ai Blog. URL: https://fireflies.ai/blog/ai-hallucinations#:~:text=GPT.

Labadze, L., Grigolia, M., & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. *International Journal of Educational Technology in Higher Education*, 20(1): 1–17. Doi: https://doi.org/10.1186/s41239-023-00426-1.

Lakhani, K. (2023). How Can We Counteract Generative AI's Hallucinations? | Digital Data Design Institute at Harvard. *Digital Data Design Institute at Harvard*. URL: https://d3.harvard.edu/how-can-we-counteract-generative-ais-hallucinations/.

Lateral reading: College students learn to critically evaluate internet sources in an online course. (2021). *Misinformation Review*. URL: https://misinforeview.hks.harvard.edu/article/lateral-reading-college-students-learn-to-critically-evaluate-internet-sources-in-an-online-course/#:~:text=to%20vet%20websites%20using%20fact,reading%20more%20often%20post%20intervention.

Lelièvre, M., Waldock, A., Liu, M., Aspillaga, N.V., Mackintosh, A., Portela, M.J.O., ... & Garrod, O.G. (2025). Benchmarking the Pedagogical Knowledge of Large Language Models. *Cornell University*. URL: https://arxiv.org/abs/2506.18710.

Lund, B.D., Lee, T.H., Mannuru, N.R., & Arutla, N. (2025). AI and Academic Integrity: Exploring Student Perceptions and Implications for Higher Education. *Journal of Academic Ethics*, 23: 1545–1565. Doi: https://doi.org/10.1007/s10805-025-09613-3.

Milovanovic, M. (2025). Solving the Very-Real Problem of AI Hallucination. *Knostic*. URL: https://www.knostic.ai/blog/ai-hallucinations.

Moore-Colyer, R. (2025). AI hallucinates more frequently as it gets more advanced – is there any way to stop it from happening, and should we even try? *LiveScience*. URL: https://www.livescience.com/technology/artificial-intelligence/ai-hallucinates-more-frequently-as-it-gets-more-advanced-is-there-any-way-to-stop-it-from-happening-and-should-we-even-try.

Nguyen, N. (2025). What is the EU AI Act? A compr-?hensive overview. FeedbackFruits: Scale high-quality learning design effortlessly. *FeedBackFruits*. URL: https://feedbackfruits.com/blog/from-regulation-to-innovation-what-the-eu-ai-act-means-for-edtech.

Regulation (EU) 2024/1689 of the european parliament and of the council of 13 June 2024. *Official Journal of the European Union*. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689.

Wedding, P. (2023). Texas A&M University-Commerce investigating after AI chatbot claims to have written every essay in professor's class. *WFAA*. URL: https://www.wfaa.com/article/news/education/texas-am-university-commerce-investigating-incident-ai-chatbot-chatgpt-claims-have-written-every-essay-in-professors-class/287-de386e17-3684-4f8c-86e1-e9d8e6f5a316.

When AI Gets It Wrong (2025): Addressing AI Hallu-?inations and Bias. *MIT Management*. URL: https://mitsloanedtech.mit.edu/ai/basics/addressing-ai-hallucinations-and-bias/.

Zhai, C., Wibowo, S., & Li, L.D. (2024). The effects of over-reliance on AI dialogue systems on students' cognitive abilities: a systematic review. *Smart Learning Environments*, 11(1): 28. Doi: https://doi.org/10.1186/s40561-024-00316-7.

MELNYK Serhii
postgraduate,
The Bohdan Khmelnytsky National University of Cherkasy

PROBLEMS OF DATA UNRELIABILITY WHEN USING ARTIFICIAL INTELLIGENCE IN EDUCATIONAL ACTIVITIES

Summary. *Problem (Introduction). The rapid integration of generative artificial intelligence (AI) into education has created unprecedented opportunities for personalized learning, yet it has also raised serious concerns about the reliability of AI-generated content. Large language models (LLMs) optimize for plausibility rather than truth, and they can fabricate facts, citations or even legal cases (When AI Gets It Wrong: Addressing AI Hallucinations and Bias - MIT Sloan Teaching & Learning Technologies, n.d.). Such hallucinations threaten academic integrity: students may unknowingly absorb falsehoods, while teachers could inadvertently reproduce inaccuracies in course materials. Empirical studies reveal that AI systems can hallucinate from less than 1 % up to 15–40 % of cases in educational tasks depending on the model and domain (Figure 1), and systematic reviews note that over-reliance on AI dialogue systems is linked to diminished critical thinking, increased technology dependence and the spread of misinformation (Zhai et al., 2024).*

Purpose. This article aims to analyze the scope and causes of AI-generated misinformation in education and to develop evidence-based recommendations for mitigating these risks. It combines technical insights on model architecture and training data with pedagogical strategies to foster AI literacy. The goal is to ensure that AI enhances rather than undermines learning.

Methods. A systematic literature review of over 80 sources, including scientific articles, policy documents (AI Act, UNESCO guidelines), and empirical studies, provided a theoretical foundation. Comparative analysis of hallucination rates across models (Makhno et al., 2025; Lelièvre et al., 2025) informed the quantitative assessment. The study also modeled mitigation strategies such as Retrieval-Augmented Generation (RAG) and Chain-of-Verification and evaluated pedagogical interventions like lateral reading and AI literacy programmes.

Results. The findings show that hallucinations stem from both internal (model architecture and context limitations) and external (biased or incomplete training data) factors. Even top models misinform 1–3% of the time, whereas widely used free systems can err 15–40% of the time when generating bibliographies or research proposals (Balch & Blanck, 2024). Hallucinations manifest in various forms: logical errors, mathematical mistakes, fabricated sources and factual inaccuracies. Their educational consequences include decreased critical thinking, increased plagiarism ("AI-giarism") and a risk of spreading disinformation. Regulatory frameworks classify educational AI systems as high-risk; Annex III of the EU AI Act lists educational AI systems for admissions, assessment and monitoring as high-risk and sets obligations for accuracy, transparency and human oversight (Nguyen, 2025). Among mitigation strategies, RAG reduces hallucinations, while Chain-of-Verification and self-consistency improve reliability. Pedagogically, teaching students lateral reading, updating academic policies, and redesigning assessments to require reflection and verification are essential.

Originality. Unlike purely technical surveys or broad commentaries, this study bridges AI research with educational practice and policy, providing a holistic perspective tailored to Ukrainian higher education. It synthesises international findings with local realities, offers a taxonomy of AI errors, presents original visualisations (Table 1 and Figure 1), and proposes a multi-level framework combining technical, pedagogical and regulatory solutions. The article emphasises that AI hallucinations are not simply technical bugs but systemic challenges requiring cultural change.

Conclusion. To harness AI's benefits in education, stakeholders must recognize and mitigate the problem of misinformation. Improving models (via RAG, verification chains), enhancing AI literacy, and adhering to high-risk regulatory standards will help ensure that AI supports, rather than sabotages, learning. Future research should focus on domain-specific hallucination rates, real-time fact-checkers for Ukrainian-language content, and longitudinal studies on AI's cognitive impact. Ultimately, balancing technological innovation with human oversight and ethical principles will determine whether AI becomes a trustworthy educational ally or a source of confusion.

Keywords*: hallucinations; artificial intelligence; data reliability; education; critical thinking; large language models; LLM; academic integrity; data verification.*

Одержано редакцією 22.08.2025
Прийнято до публікації 10.09.2025