

Пирог М.В.<https://orcid.org/0000-0003-2588-6066>

Київський національний університет імені Тараса Шевченка

Терещук Г.М.<https://orcid.org/0000-0001-7573-9748>

Київський національний університет імені Тараса Шевченка

ВАЛІДНІСТЬ ТЕСТОВОГО ОЦІНЮВАННЯ ЗНАТЬ СТУДЕНТІВ В ЕПОХУ ГЕНЕРАТИВНОГО ШТУЧНОГО ІНТЕЛЕКТУ: СИСТЕМАТИЧНИЙ ОГЛЯД

Стаття присвячена актуальній проблемі впливу генеративного штучного інтелекту на валідність тестового оцінювання знань студентів у закладах вищої освіти. Поява та швидке поширення моделей типу GPT створили серйозний виклик традиційним методам контролю знань, оскільки сучасні генеративні моделі демонструють високу ефективність у проходженні академічних тестів, особливо завдань репродуктивного характеру.

На основі систематичного огляду наукових публікацій проаналізовано ключові аспекти використання генеративних моделей ШІ в різних типах тестів, а також сучасні підходи до трансформації систем оцінювання.

Дослідження виявило, що студенти широко застосовують генеративний ШІ під час підготовки та проходження тестів, особливо асинхронних онлайн-тестів. Водночас автоматичні детектори тексту та системи онлайн-прокторингу мають суттєві обмеження щодо точності та надійності, що робить їх недостатньо ефективними для забезпечення академічної доброчесності.

У статті обґрунтовано необхідність радикального переосмислення підходів до тестового оцінювання. Запропоновано принципи AI-резистентного дизайну завдань, які базуються на підвищенні когнітивної складності, персоналізації та контекстуалізації завдань. Систематизовано альтернативні методи оцінювання.

На основі проведеного аналізу сформульовано практичні рекомендації для закладів вищої освіти України щодо короткострокової, середньострокової та довгострокової трансформації систем тестового оцінювання в умовах широкої доступності генеративного штучного інтелекту. Особлива увага приділена необхідності переходу від політики заборони до стратегії етичної та педагогічно обґрунтованої інтеграції ШІ в освітній процес.

Ключові слова: генеративний штучний інтелект, тестове оцінювання, академічна доброчесність, валідність оцінювання, детекція ШІ, AI-резистентне оцінювання, вища освіта, онлайн-тестування, автентичне оцінювання.

Постановка проблеми. Тестування залишається одним із найпоширеніших методів оцінювання знань студентів у системі вищої освіти завдяки своїй ефективності, відносній об'єктивності та можливості оперативної перевірки результатів великої кількості здобувачів освіти [1]. Проте поява генеративних моделей штучного інтелекту, що стали широко доступними для студентів і продемонстрували здатність успішно виконувати різні типи тестових завдань, поставило під сумнів валідність традиційних форм контролю знань [2].

Сучасні дослідження свідчать, що генеративні моделі ШІ демонструють високі результати у тестах, особливо у форматах де переважають завдання репродуктивного характеру [3, 4]. Водночас поширення ШІ у вищій освіті супроводжується активним використанням цих інструментів студентами для підготовки до тестування [5, 6]. Особливо вразливими виявляються асинхронні онлайн-тести, у межах яких студенти мають доступ до зовнішніх ресурсів і мінімальний рівень безпосереднього контролю.

За цих умов для закладів вищої освіти постає фундаментальна проблема: як забезпечити валід-

ність і надійність тестового оцінювання? Дослідження показують, що традиційні методи виявлення академічної нечесності мають суттєві обмеження: автоматичні детектори ШІ-генерованого контенту демонструють нестабільну точність, а редагування або перефразування суттєво знижує ефективність таких інструментів [7, 8]. Водночас заборона використання ШІ видається малореалістичною, оскільки ці технології поступово стають частиною сучасного професійного.

Для України ця проблема набуває особливої актуальності в умовах цифровізації освіти та поширення дистанційних і змішаних форматів навчання. У зв'язку з цим заклади вищої освіти потребують науково обґрунтованих підходів до адаптації систем тестового оцінювання, які б забезпечували академічну доброчесність, збереження освітніх стандартів і справедливості оцінювання.

Таким чином, виникає потреба у комплексному аналізі впливу генеративного штучного інтелекту на валідність тестування як методу оцінювання.

Аналіз останніх досліджень і публікацій.

Методологія огляду. Для систематичного аналізу проблеми впливу генеративного штучного інтелекту на тестове оцінювання студентів було проведено пошук наукових публікацій у базах даних Scopus та Web of Science. Часовий діапазон пошуку охоплював період з 2020 по 2025 рік, з особливим фокусом на публікації після листопада 2022 року – моменту публічного запуску ChatGPT. Ключовими термінами пошуку були: «ChatGPT», «generative AI», «student assessment», «testing», «multiple choice questions», «academic integrity», «AI detection», «online proctoring». Було проаналізовано 24 наукові публікації, що включають емпіричні дослідження, систематичні огляди, мета-аналізи та методичні рекомендації провідних освітніх організацій.

Тематичний аналіз досліджень. Фундаментальні дослідження впливу генеративного штучного інтелекту на освіту підкреслюють необхідність переосмислення традиційних підходів до оцінювання та забезпечення академічної доброчесності [2]. Систематичні огляди останніх років демонструють стрімке зростання інтересу до цієї проблематики та неоднозначність впливу генеративного ШІ на результати навчання [9, 10, 11].

Окремий напрям досліджень присвячений продуктивності генеративних моделей у тестових завданнях. Результати свідчать, що сучасні моделі здатні успішно виконувати завдання різних типів [3, 4]. Водночас виявляються обмеження,

пов'язані з інтерпретацією контекстуально залежних і культурно специфічних завдань [12].

Значна увага приділяється проблемі академічної доброчесності. Дослідження свідчать про широке використання інструментів ШІ у навчальному процесі, що поєднується з етичними дилемами та ризиками недоброчесної поведінки [1, 6, 13]. Водночас технологічні підходи до детекції ШІ-генерованого контенту мають обмежену ефективність [7, 8].

У зв'язку з цим сучасні дослідження фокусуються на трансформації дизайну оцінювання. Запропоновані підходи включають інтеграцію ШІ як навчального інструменту, розробку завдань вищих когнітивних рівнів, а також впровадження автентичного та процесуального оцінювання [14, 15, 16, 17]. Важливу роль відіграють також інституційні політики, спрямовані на формування етичного та відповідального використання ШІ в освіті [18].

Сучасний науковий дискурс засвідчує перехід від спроб технологічного контролю до комплексного переосмислення підходів до оцінювання.

Постановка завдання. Мета дослідження полягає у систематизації та критичному аналізі наукових досліджень щодо впливу генеративного штучного інтелекту на валідність та надійність тестування як методу оцінювання знань студентів, а також визначенні ефективних стратегій адаптації тестових систем до нових технологічних реалій.

Для досягнення поставленої мети необхідно вирішити наступні завдання:

1. Проаналізувати здатність сучасних генеративних моделей ШІ до успішного проходження різних типів академічних тестів та визначити фактори, що впливають на їхню продуктивність.
2. Систематизувати дані про використання ШІ студентами під час тестування, визначити найбільш вразливі форми тестового оцінювання.
3. Оцінити ефективність існуючих методів детекції використання ШІ в тестових відповідях.
4. Узагальнити сучасні підходи до створення AI-резистентних тестів та альтернативних методів оцінювання.
5. Сформулювати науково обґрунтовані рекомендації для закладів вищої освіти України щодо трансформації систем тестового оцінювання з урахуванням викликів епохи генеративного штучного інтелекту.

Об'єкт дослідження: процес оцінювання знань студентів закладів вищої освіти методом тестування в умовах широкої доступності генеративного штучного інтелекту.

Предмет дослідження: валідність і надійність тестового оцінювання знань студентів, а також підходи до його адаптації в умовах поширення генеративного штучного інтелекту.

Систематичний огляд наукової літератури дозволить виявити найбільш ефективні стратегії збереження академічної доброчесності при тестовому оцінюванні, окреслити перспективні напрямки трансформації систем оцінювання у вищій освіті.

Виклад основного матеріалу.

Продуктивність генеративного штучного інтелекту в проходженні тестів. Одним із ключових чинників, що впливає на валідність тестового оцінювання в епоху генеративного штучного інтелекту, є здатність сучасних моделей успішно виконувати академічні тестові завдання. Дослідження демонструють суттєву еволюцію продуктивності моделей: якщо GPT-3.5 показував обмежені результати, то GPT-4 та новіші версії досягають рівня, співставного або навіть вищого за середні показники студентів [3, 4].

Результативність генеративного ШІ значною мірою залежить від характеристик завдань. Найбільш вразливими є завдання репродуктивного характеру тоді як відкриті завдання з розгорнутими відповідями демонструють вищу стійкість до автоматизованого виконання [3, 4]. Відповідно до таксономії Блума, завдання нижчих когнітивних рівнів є більш доступними для ШІ, тоді як завдання, що передбачають аналіз, синтез, оцінювання або створення, залишаються складнішими для моделей.

Важливим фактором є також предметна специфіка та контекст завдання. Дослідження показують, що генеративні моделі демонструють вищу ефективність у технічних і формалізованих дисциплінах, тоді як у гуманітарних галузях їхня продуктивність є нижчою [12, 19]. Крім того, більшість досліджень базується на англомовних тестах, тоді як для інших мов, зокрема української, точність моделей може бути нижчою.

Попри високі результати, генеративні моделі демонструють низку обмежень. Вони можуть допускати помилки у завданнях, що вимагають глибокого контекстуального розуміння, знань «здорового глузду» або роботи з неявною інформацією, а також схильні до генерації правдоподібних, але некоректних відповідей («галюцинацій») [3]. Нижча ефективність також спостерігається у випадках використання візуальних елементів та завдань, що потребують актуальної інформації [12].

Таким чином, сучасні генеративні моделі вже становлять суттєвий виклик для традиційного

тестування, однак їхня ефективність залишається нерівномірною та залежною від структури, складності та контексту завдань.

Отримані результати свідчать, що зростання продуктивності генеративного ШІ безпосередньо впливає на поведінку студентів та практики його використання під час тестування.

Використання штучного інтелекту студентами під час тестування. Генеративний ШІ швидко інтегрувався у навчальну практику студентів і використовується для пошуку інформації, пояснення складних тем, підготовки до тестування та перевірки відповідей [5, 6]. Водночас таке використання супроводжується ризиками академічної недоброчесності та етичними дилемами [1, 18]. Аналіз досліджень дає змогу виокремити кілька моделей використання ШІ студентами: пряму генерацію відповідей, часткове використання як допоміжного інструменту для пояснення матеріалу чи структурування відповідей, а також верифікацію й перефразування відповідей з метою уникнення детекції [6, 13].

Використання ШІ студентами не слід розглядати виключно як прояв академічної недоброчесності, оскільки воно відображає ширші трансформації освітнього середовища, пов'язані з цифровізацією [6, 18]. Це зумовлює необхідність переходу від суто контрольних механізмів до переосмислення підходів до оцінювання та аналізу ефективності методів детекції.

Методи детекції використання штучного інтелекту. Зростання використання генеративного штучного інтелекту студентами стимулювало розвиток різних підходів до виявлення ШІ-генерованого контенту. Автоматичні детектори тексту демонструють нестабільну точність і залежать від характеристик тексту [7, 8].

Для тестів із множинним вибором застосовуються статистичні методи аналізу відповідей, що дозволяють виявляти аномальні патерни, зокрема невідповідність між складністю завдань і правильністю відповідей або нетипові часові характеристики виконання [1]. Однак такі підходи не забезпечують однозначного підтвердження використання ШІ, а лише ідентифікують потенційно підозрілі випадки.

Онлайн-прокторинг розглядається як додатковий інструмент контролю в умовах дистанційного оцінювання, проте його ефективність є обмеженою, а впровадження супроводжується проблемами приватності, психологічного комфорту та нерівності доступу до технологій [20, 21, 22]. Навіть у поєднанні з іншими підходами прокто-

ринг не гарантує повного запобігання використанню генеративного ШІ.

Загалом, сучасні дослідження демонструють, що детекція не може розглядатися як самодостатнє рішення проблеми академічної доброчесності. Її ефективність є обмеженою як з технічної, так і з етичної точки зору, що обумовлює необхідність переходу від контролю до переосмислення дизайну оцінювання [7, 8]. Це обумовлює необхідність переходу до педагогічно орієнтованих підходів, що розглядаються в наступному розділі.

Трансформація дизайну тестів та альтернативні методи оцінювання. Усвідомлення обмежень детекції стимулювало розробку нових підходів до оцінювання, які є більш резистентними до використання ШІ або навіть інтегрують його як легітимний інструмент навчання.

Підходи до проєктування AI-резистентного оцінювання. У роботі Майка Перкінса [14] запропоновано шкалу оцінювання штучного інтелекту (Artificial Intelligence Assessment Scale, AIAS), що являє собою комплексний підхід до етичної інтеграції генеративного ШІ в освітнє оцінювання. У межах цього підходу завдання класифікуються за рівнем допустимості використання ШІ – від повної заборони до його цілеспрямованого використання як навчального інструменту. Ключова ідея полягає в тому, що ефективно оцінювання в умовах поширення ШІ потребує не універсальних заборон, а чіткого визначення ролі технологій у кожному конкретному завданні.

У дослідженні Я Пін Сяо [15] розроблено підхід до редизайну оцінювання письмових робіт у контексті використання великих мовних моделей, який акцентує увагу на трьох ключових принципах: автентичності, персоналізації та процесуальності. Відповідно до цього підходу, завдання мають відображати реальні або професійно орієнтовані ситуації, враховувати індивідуальний досвід студентів і передбачати оцінювання не лише кінцевого результату, а й процесу його досягнення.

У роботі Тея МакЕдвардса [23] запропоновано узагальнену концепцію оцінювання з урахуванням ШІ, яка охоплює такі виміри, як педагогічна валідність, розвиток технологічної грамотності, етична рефлексія та забезпечення справедливості оцінювання. Такий підхід підкреслює необхідність інтеграції педагогічних і технологічних аспектів у процесі трансформації систем оцінювання.

Принципи AI-резистентного дизайну тестів. Сучасні дослідження демонструють, що ефективність оцінювання в умовах поширення генера-

тивного ШІ значною мірою залежить від дизайну завдань.

Зокрема, у дослідженні Бін Нгуен Тханя [16] емпірично доведено, що завдання, орієнтовані на вищі рівні когнітивної складності (аналіз, синтез, оцінювання, створення), є більш стійкими до використання ШІ порівняно із завданнями репродуктивного характеру.

Одним із ключових принципів є когнітивна складність вищого порядку. Завдання, що передбачають порівняння, аргументацію, оцінювання альтернатив або прийняття рішень у нових контекстах, суттєво ускладнюють можливість прямого використання згенерованих відповідей.

Важливим є також принцип контекстуалізації та персоналізації. У роботі Я Пін Сяо [15] підкреслено, що завдання, які передбачають інтеграцію особистого досвіду студента, аналіз локальних кейсів або власних навчальних проєктів, є менш вразливими до використання ШІ, оскільки базуються на унікальному контексті.

Ще одним напрямом є мультимодальність завдань, що передбачає використання не лише текстових, а й візуальних, аудіальних або інтерактивних компонентів. Такий підхід підвищує складність автоматизованого генерування відповідей і сприяє більш комплексному оцінюванню знань.

Крім того, важливим є впровадження процесуального оцінювання, яке передбачає аналіз не лише кінцевого результату, але й процесу його створення. Це може включати поетапне виконання завдань, подання чернеток, пояснення логіки розв'язання або усний захист результатів. Такий підхід суттєво знижує можливості неконтрольованого використання генеративного ШІ.

Узагальнення результатів аналізу дозволяє систематизувати основні AI-резистентні підходи до проєктування оцінювання, які наведено в табл. 1.

Представлені підходи відображають перехід від технологічного контролю до педагогічного редизайну оцінювання, орієнтованого на розвиток складних когнітивних навичок та автентичного навчання.

Специфічні стратегії для різних типів тестів. Підвищення стійкості тестів з множинним вибором може досягатися через використання ситуаційних завдань, елементів інтерпретації та варіантів відповідей, що перевіряють глибину розуміння, а не лише відтворення знань [16]. Відкриті завдання з обґрунтуванням логіки відповіді є більш валідними в умовах поширення ШІ, оскільки ускладнюють використання готових згенерованих рішень [15]. Додатковим обмежуваль-

Систематизація AI-резистентних підходів до оцінювання у вищій освіті

Підхід	Сутність	Чому резистентний до ШІ	Приклад реалізації	Джерела
Завдання вищих когнітивних рівнів	Орієнтація на аналіз, синтез, оцінювання та створення	Вимагають критичного мислення, аргументації та нестандартних рішень	Порівняння підходів, обґрунтування вибору, кейс-аналіз	Nguyen Thanh et al., 2023
Контекстуалізовані та персоналізовані завдання	Використання індивідуального досвіду, локальних кейсів	ШІ не має доступу до особистого досвіду студента	Аналіз власного проєкту, практики або досвіду	Hsiao et al., 2023
Процесуальне оцінювання	Оцінка не лише результату, а й процесу виконання	Ускладнює повне делегування роботи ШІ	Чернетки, пояснення кроків, усний захист	Hsiao et al., 2023; Sullivan et al., 2023
Автентичне оцінювання	Завдання, наближені до реальних професійних ситуацій	Потребують інтеграції знань і прийняття рішень у складних умовах	Бізнес-кейс, реальна задача з галузі	Perkins et al., 2024
Мультимодальні завдання	Поєднання тексту, графіків, відео, даних	Знижують ефективність текстових моделей	Аналіз діаграми + текстове пояснення	Beltozar-Clemente et al., 2024
Відкриті питання з поясненням логіки	Вимога обґрунтування відповіді	ШІ може генерувати відповідь, але складніше пояснити індивідуальну логіку	“Поясніть, чому обрано саме цей підхід”	Hsiao et al., 2023
Синхронне тестування з обмеженням часу	Часові обмеження та контроль умов	Зменшує можливість використання зовнішніх ресурсів	Онлайн-тест із таймером	Newton & Essex, 2023
Інтеграція ШІ як інструменту	Дозволене використання ШІ з критичним аналізом	Зміщує фокус з контролю на мислення	Проаналізувати відповідь ChatGPT і знайти помилки	Perkins et al., 2024; Oates & Johnson, 2025
Портфоліо з рефлексією	Оцінка довготривалого навчального процесу	Важко повністю автоматизувати	Збір робіт + самоаналіз	Perkins et al., 2024
Усне оцінювання	Діалогова перевірка знань	Вимагає спонтанного мислення	Усний іспит або презентація	Oates & Johnson, 2025

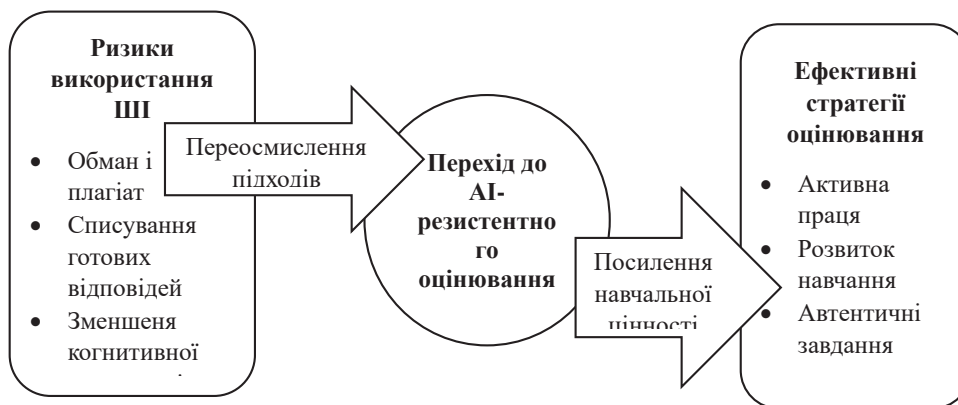


Рис. 1. Концептуальна модель AI-резистентного оцінювання

ним чинником є синхронне тестування з чіткими часовими межами, хоча воно не усуває проблему повністю [1].

Альтернативні методи оцінювання. Дослідження демонструють необхідність диверсифікації методів оцінювання у відповідь на виклики генеративного ШІ. Найбільш ефективними є підходи, що орієнтуються на процес навчання та активну участь студента, зокрема портфоліо

з рефлексією [14], усні форми оцінювання [17], проєктне оцінювання з поетапним контролем [24] та колаборативні формати. Ці методи ускладнюють повну автоматизацію виконання завдань за допомогою ШІ та сприяють розвитку критичного мислення й глибшого розуміння матеріалу.

Інтеграція ШІ як навчального інструменту. Перспективним напрямом є не заборона, а контрольована інтеграція генеративного штучного інте-

лекту в навчальний процес. У дослідженні Майка Перкінса [14] підкреслено, що навчальні завдання можуть передбачати використання ШІ з подальшим критичним аналізом отриманих результатів, їх перевіркою та адаптацією. Такий підхід сприяє розвитку критичного мислення та формує відповідальне ставлення до використання технологій.

Виклики впровадження нових підходів. Попри теоретичну обґрунтованість AI-резистентних методів оцінювання, їх практична реалізація супроводжується низкою суттєвих викликів. У роботі Міріам Салліван [24] зазначено, що альтернативні підходи потребують значно більших часових і організаційних ресурсів з боку викладачів, зокрема для розробки складніших завдань, надання індивідуалізованого зворотного зв'язку та багатокритеріального оцінювання результатів. В умовах масового навчання та обмежених ресурсів це ускладнює їх масштабування.

У дослідженні Бін Нгуєн Тханя [16] також підкреслено, що не всі навчальні результати можуть бути ефективно оцінені за допомогою альтернативних методів. Зокрема, оцінювання базових знань і фундаментальних навичок у багатьох випадках і надалі потребує використання тестових форм контролю, які, однак, мають бути адаптовані з урахуванням можливостей генеративного штучного інтелекту.

Водночас сучасні дослідження вказують на те, що трансформація оцінювання не може обмежуватися лише зміною інструментів або форматів завдань. Ефективне впровадження нових підходів передбачає комплексні зміни, включаючи професійний розвиток викладачів, перегляд освітніх практик і формування готовності до експериментування з інноваційними моделями оцінювання [14, 18].

Інституційні політики та педагогічні стратегії. Ефективна відповідь на виклики генеративного штучного інтелекту потребує системних змін на рівні освітніх інституцій і педагогічних підходів [2]. Відсутність чітких політик призводить до неоднозначних практик використання ШІ, тоді як інституційне лідерство відіграє ключову роль у формуванні етичних підходів [18].

У науковій літературі виокремлюють три основні підходи до регулювання використання ШІ: рестриктивний, регульований та інтеграційний, який передбачає використання ШІ як навчального інструменту і вважається найбільш перспективним [14, 18].

Ефективна інституційна політика має включати чіткі правила використання ШІ, освітній компонент щодо його етичного застосування, під-

тримку викладачів у трансформації оцінювання та врахування специфіки дисциплін [2, 14, 18].

Педагогічні стратегії відіграють ключову роль у забезпеченні академічної доброчесності. Зокрема, формування внутрішньої мотивації студентів [18], зниження навчального стресу та розвиток саморегуляції [6], розвиток критичного мислення у взаємодії з ШІ [17], а також впровадження формуючого оцінювання [24] сприяють більш відповідальному використанню технологій.

Водночас впровадження нових підходів супроводжується низкою викликів, зокрема підвищеним навантаженням на викладачів, проблемами масштабованості, опором змінам та нерівністю доступу до цифрових ресурсів [2, 18, 24].

Таким чином, трансформація оцінювання в умовах поширення генеративного ШІ має базуватися на комплексному поєднанні інституційних, педагогічних і технологічних рішень, із фокусом на інтеграцію ШІ як інструменту розвитку критичного мислення та відповідального використання технологій [14, 17].

Висновки. Сучасні генеративні моделі демонструють високі результати у тестах, що ставить під сумнів валідність традиційного тестування [3, 4]. Студенти активно використовують ШІ у навчанні, що поєднується з ризиками академічної недоброчесності [1, 6], тоді як існуючі методи детекції залишаються обмежено ефективними [7, 21]. Найбільш перспективним напрямом є трансформація оцінювання через впровадження завдань вищого когнітивного рівня та альтернативних форматів [14–16]. Ефективна стратегія полягає у переході від заборони до інтеграції ШІ як інструменту розвитку критичного мислення [17].

Практичні рекомендації для закладів вищої освіти України. Короткострокові заходи включають розробку інституційної політики щодо використання ШІ, підготовку викладачів і студентів до його етичного застосування, а також частковий перехід до синхронного тестування та диверсифікації методів оцінювання.

Середньострокові заходи передбачають створення банків AI-резистентних завдань, впровадження автентичного та формуючого оцінювання, а також інтеграцію усних форм контролю й завдань із критичним використанням ШІ.

Довгострокова перспектива полягає у пересмисленні цілей оцінювання з акцентом на розвиток критичного мислення, креативності та етичної рефлексії, а також у системній інтеграції AI-грамотності й підтримці викладачів у впровадженні інноваційних підходів.

Напрямки майбутніх досліджень. Систематичний огляд виявив низку ключових прогалин, що можуть бути згруповані у такі напрями.

1. Контекстуальні та емпіричні дослідження. Потребують розвитку дослідження в українському освітньому контексті специфіки використання ШІ студентами та викладачами, аспектів академічної доброчесності, а також дослідження довгострокового впливу трансформації оцінювання на якість навчання та професійну підготовку випускників.

2. Оцінка ефективності та розробка інструментів. Актуальними є порівняльні експериментальні дослідження ефективності різних стратегій AI-резистентного оцінювання, а також розробка

й емпірична валідація нових інструментів оцінювання.

3. Педагогічні та етичні моделі інтеграції ШІ. Подальших досліджень потребує визначення оптимального балансу між контролем і автентичністю оцінювання, а також розробка ефективних педагогічних моделей інтеграції ШІ освітній процес.

Генеративний штучний інтелект не означає кінець тестування, але вимагає його суттєвої трансформації. Перспективним є підхід, у якому ШІ інтегрується як інструмент навчання, а оцінювання фокусується на розвитку критичного мислення, креативності та здатності до безперервного навчання.

Список літератури:

1. Newton P. M., Essex K. Cheating in exams: A comprehensive review of data from 19 studies. *Assessment & Evaluation in Higher Education*. 2023. Vol. 48, No 8. С. 1105–1119. DOI: <https://doi.org/10.1080/02602938.2022.2161391>

2. Dwivedi Y. K., Kshetri N., Hughes L., Slade E. L., Jeyaraj A., Kar A. K. та ін. "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*. 2023. Vol. 71. 102642. DOI: <https://doi.org/10.1016/j.ijinfomgt.2023.102642>

3. Newton P., Xiromeriti M. ChatGPT performance on multiple choice question examinations in higher education. A pragmatic scoping review. *Assessment & Evaluation in Higher Education*. 2024. Vol. 49, No 6. С. 781–798. DOI: <https://doi.org/10.1080/02602938.2023.2299059>

4. Giunti M., Garavaglia F. G., Giuntini R., Sergioli G., Pinna S. ChatGPT as a prospective undergraduate and medical school student. *PLoS ONE*. 2024. Vol. 19, No 10. e0308157. DOI: <https://doi.org/10.1371/journal.pone.0308157>

5. Baidoo-Anu D., Ansah L. O. Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI*. 2023. Vol. 7, No 1. С. 52–62. DOI: <https://doi.org/10.61969/jai.1337500>

6. Uppal K., Hajian S. Students' perceptions of ChatGPT in higher education: A study of academic enhancement, procrastination, and ethical concerns. *European Journal of Educational Research*. 2025. Vol. 14, No 1. С. 199–211. DOI: <https://doi.org/10.12973/eu-jer.14.1.199>

7. Malik M. A., Amjad A. I. AI vs AI: How effective are Turnitin, ZeroGPT, GPTZero, and Writer AI in detecting text generated by ChatGPT, Perplexity, and Gemini? *Journal of Applied Learning & Teaching*. 2025. Vol. 8, No 1. С. 45–58. DOI: <https://doi.org/10.37074/jalt.2025.8.1.9>

8. Susnjak T. Maintaining academic integrity in the age of ChatGPT: Challenges and opportunities for online assessments. *Assessment & Evaluation in Higher Education*. 2024. Vol. 49, No 4. С. 456–472. DOI: <https://doi.org/10.1080/02602938.2023.2289089>

9. Yan L. Eight months into reality: A scoping review of the application of ChatGPT in higher education teaching and learning. *Innovative Higher Education*. 2023. Vol. 49, No 2. С. 245–268. DOI: <https://doi.org/10.1007/s10755-025-09790-4>

10. Mai D. T. T., Da C. V., Hanh N. V. The use of ChatGPT in teaching and learning: A systematic review through SWOT analysis approach. *Frontiers in Education*. 2024. Vol. 9. 1328769. DOI: <https://doi.org/10.3389/educ.2024.1328769>

11. Deng R., Jiang M., Yu X., Lu Y., Liu S. Does ChatGPT enhance student learning? A systematic review and meta-analysis of experimental studies. *Computers & Education*. 2024. 105224. DOI: <https://doi.org/10.1016/j.compedu.2024.105224>

12. Beltozar-Clemente S., Díaz-Vega E., Zapata-Paulini J., Tejada-Navarrete R. E. We Can Rely on ChatGPT as an Educational Tutor: A Cross-Sectional Study of its Performance, Accuracy, and Limitations in University Admission Tests. *International Journal of Engineering Pedagogy (iJEP)*. 2024. Vol. 14, No 1. С. 50–60. DOI: <https://doi.org/10.3991/ijep.v14i1.46787>

13. Cotton D. R. E., Cotton P. A., Shipway J. R. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*. 2023. Vol. 61, No 2. С. 228–239. DOI: <https://doi.org/10.1080/14703297.2023.2190148>

14. Perkins M., Furze L., Roe J., MacVaugh J. The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice*. 2024. Vol. 21, No 06. Article 03. DOI: <https://doi.org/10.53761/q3azde36>
15. Hsiao Y. P., Klijn N., Chiu M. S. Developing a framework to re-design writing assignment assessment for the era of Large Language Models. *Learning: Research and Practice*. 2023. Vol. 9, No 2. C. 148–158. DOI: <https://doi.org/10.1080/23735082.2023.2257234>
16. Nguyen Thanh B., Vo D. T. H., Nguyen Nhat M., Pham T. T. T., Thai Trung H., Ha Xuan S. Race with the machines: Assessing the capability of generative AI in solving authentic assessments. *Australasian Journal of Educational Technology*. 2023. Vol. 39, No 5. C. 59–81. DOI: <https://doi.org/10.14742/ajet.8902>
17. Oates A., Johnson D. ChatGPT in the classroom: Evaluating its role in fostering critical evaluation skills. *International Journal of Artificial Intelligence in Education*. 2025. Vol. 35. C. 1793–1824. DOI: <https://doi.org/10.1007/s40593-024-00452-8>
18. Crawford J., Cowling M., Allen K.-A. Leadership is needed for ethical ChatGPT: Character, assessment, and learning using artificial intelligence (AI). *Journal of University Teaching & Learning Practice*. 2023. Vol. 20, No 3. Article 02. DOI: <https://doi.org/10.53761/1.20.3.02>
19. Jauhainen J. S., Garagorry Guerra A. Generative AI in education: ChatGPT-4 in evaluating students' written responses. *Innovations in Education and Teaching International*. 2025. Vol. 62, No 4. C. 1377–1394. DOI: <https://doi.org/10.1080/14703297.2024.2422337>
20. Maphalala M. C., Nkosi N. Proctoring Online Assessments: Enhancing Security and Academic Integrity in Open Distance eLearning. *Open Praxis*. 2025. Vol. 17, No 4. C. 680–690. DOI: <https://doi.org/10.55982/openpraxis.17.4.911>
21. Mutimukwe C., Viberg O., McGrath C., Cerratto-Pargman T. Privacy in online proctoring systems in higher education: stakeholders' perceptions, awareness and responsibility. *Journal of Computing in Higher Education*. 2025. DOI: <https://doi.org/10.1007/s12528-025-09461-5>
22. McEdwards T., Underhill G. R. The element of choice: Online students' perceptions of online exam proctoring. *Online Journal of Distance Learning Administration*. 2025. Vol. 28, No 1.
23. Kılınc S. Comprehensive AI assessment framework: Enhancing educational evaluation with ethical AI integration. *arXiv preprint*. 2024. DOI: <https://doi.org/10.48550/arXiv.2407.16887>
24. Sullivan M., Kelly A., McLaughlan P. ChatGPT in higher education: Considerations for academic integrity and student learning. 1. 2023. Vol. 6, no. 1. DOI: <https://doi.org/10.37074/jalt.2023.6.1.17>

Pyroh M.V., Tereshchuk H.M. THE VALIDITY OF TEST-BASED ASSESSMENT OF STUDENT KNOWLEDGE IN THE AGE OF GENERATIVE ARTIFICIAL INTELLIGENCE: A SYSTEMATIC REVIEW

This article addresses the pressing issue of how generative artificial intelligence affects the validity of student assessment in higher education institutions. The emergence and rapid spread of GPT-type models have posed a serious challenge to traditional methods of knowledge assessment, as modern generative models demonstrate high effectiveness in passing academic tests, especially tasks of a reproductive nature.

Based on a systematic review of scientific publications, this study analyzes key aspects of the use of generative AI models in various types of tests, as well as modern approaches to transforming assessment systems.

The study found that students widely use generative AI when preparing for and taking tests, especially asynchronous online tests. At the same time, automatic text detectors and online proctoring systems have significant limitations in terms of accuracy and reliability, making them insufficiently effective for ensuring academic integrity.

The article justifies the need for a radical rethinking of approaches to test-based assessment. Principles of AI-resistant task design are proposed, based on increased cognitive complexity, personalization, and contextualization of tasks. Alternative assessment methods are systematized.

Based on the analysis, practical recommendations have been formulated for higher education institutions in Ukraine regarding the short-term, medium-term, and long-term transformation of testing and assessment systems in the context of the widespread availability of generative artificial intelligence. Particular attention is paid to the need to shift from a policy of prohibition to a strategy of ethical and pedagogically sound integration of AI into the educational process.

Keywords: *generative artificial intelligence, test-based assessment, academic integrity, assessment validity, AI detection, AI-resistant assessment, higher education, online testing, authentic assessment.*

Дата першого надходження статті до видання: 17.03.2026

Дата прийняття статті до друку після рецензування: 13.04.2026

Дата публікації (оприлюднення) статті: 19.05.2026