

УДК 37.091.26:811.111:004.8

DOI <https://doi.org/10.24919/2308-4863/97-3-46>**Ратмір ТУРЧАНІНОВ,***orcid.org/0000-0003-2628-3101*

аспірант кафедри спорту та фітнесу

Чернівецького національного університету імені Юрія Федьковича

(Чернівці, Україна) *ratmirvariag@gmail.com***Анжеліна РАДИК,***orcid.org/0009-0002-7425-441X*

студентка факультету іноземних мов

Чернівецького національного університету імені Юрія Федьковича

(Чернівці, Україна) *anzhelina.radyk2@gmail.com***Артем ГУРГІШ,***orcid.org/0000-0001-7433-110X*

аспірант кафедри англійської мови

Чернівецького національного університету імені Юрія Федьковича

(Чернівці, Україна) *hurhish.artem@chnu.edu.ua*

## ОЦІНЮВАННЯ ПИСЬМОВИХ РОБІТ УЧНІВ НА УРОКАХ АНГЛІЙСЬКОЇ МОВИ ЗА ДОПОМОГОЮ ШТУЧНОГО ІНТЕЛЕКТУ GOOGLE GEMINI

У статті розглянуто проблему оцінювання письмових робіт учнів на уроках англійської мови в умовах зростаючого навантаження на вчителя та обмеженості часу, що ускладнює забезпечення об'єктивності й індивідуалізації зворотного зв'язку. Актуальність дослідження зумовлена необхідністю пошуку ефективних інструментів автоматизації освітнього процесу, зокрема використання технологій штучного інтелекту. Метою роботи є експериментальне обґрунтування доцільності застосування системи штучного інтелекту Google Gemini для оцінювання письмових робіт учнів шляхом порівняння результатів автоматизованого та традиційного (людського) оцінювання.

У дослідженні використано корпус із 44 учнівських есе, які були незалежно перевірені двома рецензентами та системою штучного інтелекту Google Gemini за єдиними критеріями оцінювання (граматика, логіка, оригінальність, приклади, цитування). Отримані результати свідчать, що середній бал, виставлений ШІ, є вищим за людський приблизно на 6,5 бала, що вказує на наявність систематичного позитивного зміщення. Водночас встановлено високий рівень узгодженості між людськими оцінювачами, що підтверджує надійність традиційного підходу. Дослідження також показало, що ШІ здатен надавати персоналізований і розгорнутий зворотний зв'язок, аргументуючи оцінки та наводячи приклади помилок, що сприяє формуванню письмової компетентності учнів.

Окремо встановлено, що використання штучного інтелекту дозволяє значно скоротити час оцінювання: повний цикл перевірки робіт займає близько 50 хвилин, тоді як людські рецензенти витрачають понад 4 години. Разом із тим виявлено певні обмеження, зокрема труднощі у зчитуванні рукописного тексту, недостатню чутливість до критичних помилок і тенденцію до завищення оцінок у слабких роботах. Узагальнені результати дослідження свідчать про доцільність використання ШІ як допоміжного інструменту оцінювання за умови поєднання його з експертною оцінкою вчителя, що забезпечує баланс між ефективністю та об'єктивністю освітнього процесу.

**Ключові слова:** штучний інтелект, оцінювання есе, Google Gemini, письмові роботи, англійська мова, автоматизоване оцінювання, зворотний зв'язок.

**Ratmir TURCHANINOV,**

*orcid.org/0000-0003-2628-3101*

*PhD student at the Department of Sports and Fitness  
Yurii Fedkovich Chernivtsi National University  
(Chernivtsi, Ukraine) ratmirvariag@gmail.com*

**Anzhelina RADYK,**

*orcid.org/0009-0002-7425-441X*

*Student at the Faculty of Foreign Languages  
Yurii Fedkovich Chernivtsi National University  
(Chernivtsi, Ukraine) anzhelina.radyk2@gmail.com*

**Artem HURHISH,**

*orcid.org/0000-0001-7433-110X*

*PhD student at the Department of English  
Yurii Fedkovich Chernivtsi National University  
(Chernivtsi, Ukraine) hurhish.artem@chnu.edu.ua*

## ASSESSING STUDENTS' ESSAYS ON ENGLISH LANGUAGE CLASSES USING ARTIFICIAL INTELLIGENCE BASED ON GOOGLE GEMINI

*The article addresses the problem of assessing students' written essays in English language classes under conditions of increased teacher workload and limited time, which complicates the provision of objective and individualized feedback. The relevance of the study is обусловлена the need to identify effective tools for automating the assessment process, particularly through the use of artificial intelligence technologies. The aim of the research is to experimentally justify the feasibility of using Google Gemini AI for essay assessment by comparing automated and traditional (human) evaluation results.*

*The study is based on a dataset of 44 student essays independently assessed by two human reviewers and Google Gemini using unified evaluation criteria, including grammar, logic, originality, examples, and citations. The findings indicate that the average score assigned by AI is approximately 6.5 points higher than that of human evaluators, revealing a systematic positive bias. At the same time, a high level of consistency between human reviewers was observed, confirming the reliability of traditional assessment practices. The study also demonstrates that AI provides detailed and personalized feedback, explaining strengths and weaknesses of each essay and supporting its scores with specific examples, thus contributing to the development of students' writing skills.*

*Additionally, the results show a significant advantage of AI in terms of efficiency, as the complete assessment process, including scanning and uploading, took about 50 minutes, compared to more than four hours required by human reviewers. However, several limitations were identified, including difficulties in processing handwritten text, reduced sensitivity to critical errors, and a tendency to overestimate weaker essays. Overall, the findings support the use of AI as a supportive tool in essay assessment, provided it is combined with human expertise to ensure both efficiency and validity in the educational process.*

**Key words:** *artificial intelligence, essay assessment, Google Gemini, written works, English language, automated scoring, feedback.*

**Постановка проблеми.** Одним із найбільших викликів у професійній діяльності сучасних вчителів залишається нестача часу через великі обсяги навантаження (Олексенко, Хоменко, 2023). Разом з цим, особливої важливості набуває проблема ефективного та об'єктивного оцінювання письмових робіт учнів, що традиційно потребує значних часових витрат і високого рівня концентрації з боку педагога. Перевірка великої кількості текстових завдань ускладнюється необхідністю одночасного врахування правильності використання лексичних норм, логічної структури, аргументованості та креативності висловлювання, що створює додаткове когнітивне навантаження на вчителя.

Ситуація ускладнюється також вимогами до індивідуалізації навчання та надання розгорнутого зворотного зв'язку кожному учневі, що в умовах обмеженого часу часто реалізується частково або формально. У результаті виникає суперечність між потребою у якісному оцінюванні та реальними можливостями педагогів забезпечити його на належному рівні.

У зв'язку з цим зростає інтерес до використання технологій штучного інтелекту як інструменту автоматизації та оптимізації процесу оцінювання письмових робіт. Проте впровадження таких технологій супроводжується низкою невирішених проблем, серед яких – забезпечення валідності та надійності оцінювання, уникнення

алгоритмічної упередженості, збереження педагогічної доцільності та дотримання принципів академічної доброчесності. Отже, постає необхідність наукового обґрунтування доцільності та умов ефективного використання штучного інтелекту в оцінюванні письмових робіт учнів.

**Мета статті.** Метою даної статті є експериментальне обґрунтування доцільності використання технологій штучного інтелекту в процесі оцінювання письмових робіт учнів на уроках англійської мови шляхом порівняльного аналізу результатів автоматизованої та традиційної (викладацької) перевірки.

Дослідження спрямоване на визначення ефективності застосування ШІ з огляду на такі ключові параметри, як швидкість оцінювання, об'єктивність виставлення балів, узгодженість результатів із експертною оцінкою вчителя, а також якість наданого зворотного зв'язку.

Для досягнення поставленої мети було використано корпус реальних письмових робіт учнів, які були незалежно перевірені двома педагогами і оброблені за допомогою інструментів штучного інтелекту. Порівняння отриманих результатів дозволило оцінити потенціал автоматизованих систем у зниженні часових витрат учителя та підвищенні стандартизації оцінювання.

Окрім цього, мета статті передбачає виявлення можливих обмежень і ризиків використання ШІ в освітньому процесі, зокрема щодо точності інтерпретації змісту учнівських текстів, здатності враховувати контекст і творчі аспекти письма, а також забезпечення педагогічної доцільності такого інструменту. Таким чином, дослідження покликане не лише визначити ефективність ШІ як інструменту оцінювання, але й окреслити умови його доцільного та відповідального впровадження в практику сучасної освіти.

**Аналіз досліджень.** Проблема використання штучного інтелекту в оцінюванні письмових робіт учнів є предметом активного наукового дискурсу в межах педагогіки, прикладної лінгвістики та комп'ютерних наук. Аналіз сучасних досліджень засвідчує поступову еволюцію підходів до автоматизованого оцінювання – від спрощених статистичних моделей до складних інтелектуальних систем, що базуються на методах обробки природної мови та глибокого машинного навчання (Manning et al., 2025).

Ранні наукові праці зосереджувалися на розробці систем автоматизованого оцінювання есе, які використовували кількісні показники тексту (довжину, частоту складних слів, синтаксичні структури) як основні предиктори якості письмо-

вої роботи (Task, 2025). Однак такі підходи зазнавали критики через їхню обмежену здатність оцінювати змістову глибину та аргументативну якість тексту (Hussein et al., 2019). Подальший розвиток досліджень пов'язаний із впровадженням методів обробки природної мови та трансформерних нейронних моделей, які дозволяють здійснювати багаторівневий аналіз тексту – від лінгвістичних до семантичних і дискурсивних характеристик (Liu et al., 2025). Застосування моделей типу BERT і GPT забезпечує більш точне оцінювання завдяки врахуванню контексту та логіки викладу.

Окремий напрям сучасних досліджень присвячений використанню великих мовних моделей у задачах оцінювання письмових робіт. Доведено, що ефективність таких систем залежить від стратегій взаємодії (zero-shot, few-shot, chain-of-thought), які впливають на валідність і узгодженість результатів. Водночас дослідники наголошують на проблемі варіативності оцінок та їх нестабільності при використанні генеративних моделей (García-Varela et al., 2025).

У сучасних працях також розглядаються підходи до підвищення надійності оцінювання, зокрема використання детермінованих параметрів моделей, розширених рубрик та алгоритмічних таблиць прийняття рішень. Поряд із технологічними аспектами, значна увага приділяється педагогічним можливостям застосування ШІ. Дослідження підтверджують, що автоматизовані системи оцінювання сприяють забезпеченню оперативного зворотного зв'язку, розвитку само-рефлексії та персоналізації навчання (Fontanilla et al., 2025). Крім того, використання ШІ позитивно впливає на розвиток письмової компетентності учнів, зокрема лексико-граматичних навичок і логічної організації тексту (Zhou et al., 2025). Численними компаніями з розробки програмного забезпечення також зазначаються значні переваги у використанні миттєвого зворотнього зв'язку під час використання штучного інтелекту в освітніх контекстах (SapientPro, 2026).

Водночас численні дослідження акцентують увагу на ризиках використання ШІ в оцінюванні. Серед них є обмежене розуміння змісту, орієнтація на формальні ознаки тексту, алгоритмічна упередженість та загрози академічній доброчесності. Окремо підкреслюється ризик формування когнітивної залежності учнів від ШІ та зниження рівня критичного мислення (Deer, Chen, 2025).

Таким чином, аналіз наукових досліджень і публікацій свідчить про значний потенціал використання штучного інтелекту в оцінюванні письмових робіт, водночас підкреслюючи наявність

суттєвих методологічних, етичних і педагогічних викликів, що потребують подальшого наукового осмислення.

**Виклад основного матеріалу.** У межах проведеного аналізу було опрацьовано результати оцінювання 44 есе учнів 8, 9, 10, 11 класів. Кожна робота оцінювалася трьома незалежними оцінювачами: двома рецензентами-людьми та системою штучного інтелекту Google Gemini. Така трикомпонентна модель дозволила здійснити порівняльний аналіз між людським та автоматизованим оцінюванням. Задля прозорості та достовірності дослідження, рецензенти на система ШІ керувалися однаковими критеріями оцінювання, де було вказано максимальну кількість балів кожного пункту з поясненнями (табл. 1).

У результаті обчислення середніх значень встановлено, що середній бал першого рецензента становить приблизно 64,2 бала, другого рецензента – 63,8 бала, тоді як середній показник Google Gemini є суттєво вищим і становить близько 70,5 бала. Узагальнений середній бал по всій вибірці дорівнює приблизно 66,2 бала. При цьому середнє значення людського оцінювання (як середнє арифметичне двох рецензентів) становить близько 64,0 бала, що дозволяє здійснити безпосереднє порівняння з результатами штучного інтелекту.

Аналіз різниці між оцінками показує наявність систематичного зсуву: оцінки, виставлені ШІ, у середньому перевищують людські приблизно на 6,5 бала. Це свідчить про наявність позитивного зміщення (bias) у бік більш високого оцінювання з боку автоматизованої системи. Відповідно, абсолютна середня похибка між людськими оцінками та оцінками ШІ також становить близько 6,5 бала, що у відносному вимірі дорівнює приблизно 10,2% від середнього людського балу. Такий рівень відхилення можна вважати статистично значущим у контексті освітнього оцінювання.

Окремої уваги заслуговує рівень узгодженості між людськими рецензентами. Дані свідчать про високий ступінь консистентності: розбіжність між Рецензентом 1 та Рецензентом 2 у більшості випадків не перевищує 0-5 балів, що підтверджує надійність людського оцінювання як еталонного. Натомість ШІ демонструє іншу модель оцінювання: він більш лояльний до мовних помилок і стилістичних недоліків, натомість приділяє більше уваги загальній логіці, структурі та наявності аргументації, навіть якщо вона реалізована частково.

Особливо показовими є крайні значення. У випадку слабких робіт (із балами нижче 30) штучний інтелект часто суттєво підвищує оцінку, інколи на 15-20 балів, що свідчить про недостатню чутливість до критичних помилок або нерозуміння завдання. Водночас для сильних есе (80+ балів) різниця між людським і машинним оцінюванням є мінімальною та зазвичай не перевищує 3 балів, що вказує на високу узгодженість у визначенні якісних робіт.

У процесі оцінювання есе система штучного інтелекту продемонструвала здатність до надання розгорнутого та персоналізованого зворотного зв'язку для кожної роботи. Коментарі не обмежувалися лише виставленням балу, а містили детальні пояснення щодо сильних сторін тексту, таких як логічна структура, аргументація або використання прикладів, а також конкретні рекомендації щодо покращення. Зокрема, ШІ вказував на недоліки у граматиці, лексиці, структурі або відповідності темі, пропонуючи учням орієнтири для подальшого розвитку письмових навичок. Такий підхід відповідає сучасним педагогічним практикам формування оцінювання, де важливим є не лише результат, але й процес навчання.

Водночас під час аналізу було виявлено, що штучний інтелект стикався з певними технічними труднощами, пов'язаними зі зчитуванням руко-

Таблиця 1

**Критерії оцінювання письмових робіт учнів**

<b>Критерій оцінювання</b>	<b>Опис та деталізація</b>	<b>Макс. бали</b>
Грамаітичні та стилістичні норми	Правильність грамаітичних конструкцій, відповідність стилю та коректність синтаксису.	20
Логічність та послідовність	Логічний розвиток думок, чітка структура (вступ, основна частина, висновок).	20
Оригінальність та креативність	Інноваційність підходу, свіжість ідей та здатність розкрити тему нестандартно.	15
Приклади з власного життя	Включення особистих історій або прикладів, що підкріплюють аргументацію.	20
Приклади з актуальних подій	Використання реальних фактів або подій, що мають значення для сучасного суспільства.	20
Цитування	Вміння навести цитату відомих особистостей, що підтверджує або розширює погляд.	5
<b>ЗАГАЛЬНИЙ БАЛ</b>		<b>100</b>

писного (курсивного) тексту. У випадках, коли почерк учня був складним для розпізнавання або частини тексту були нечіткими, система прямо повідомляла про це у своїх коментарях. Така прозорість є важливою характеристикою роботи ШІ, оскільки дозволяє відокремити об'єктивні труднощі обробки даних від власне оцінювання змісту роботи. Крім того, це свідчить про наявність метакогнітивного компоненту в алгоритмах оцінювання, коли система здатна рефлексувати обмеження власного аналізу.

Окремо слід підкреслити, що оцінювання, здійснене ШІ, характеризується високим рівнем структурованості відповідно до заданих критеріїв. Система послідовно аналізувала кожен аспект роботи (граматику, логіку, оригінальність, наявність прикладів, персонального досвіду та цитат) і надавала чітко аргументовані бали за кожен із них. При цьому ШІ часто супроводжував оцінку конкретними прикладами помилок або невдалих мовних конструкцій, що підвищує валідність і прозорість оцінювання. Такий підхід дозволяє не лише обґрунтувати виставлений бал, але й забезпечує учням зрозумілий зв'язок між їхніми помилками та підсумковим результатом, що є важливим елементом ефективного навчального процесу.

Важливим аспектом проведеного дослідження є також порівняння часових витрат на оцінювання робіт. Згідно з отриманими даними, Рецензент 1 витратив на перевірку всіх есе 4 години 35 хвилин, тоді як Рецензент 2–4 години 10 хвилин. Натомість система штучного інтелекту, включно з етапами сканування та завантаження робіт, виконала повний цикл оцінювання за приблизно 50 хвилин. Таким чином, використання ШІ дозволило скоротити час перевірки більш ніж у п'ять разів порівняно з людськими рецензентами.

**Висновки.** Ці результати свідчать про суттєву перевагу автоматизованого підходу з точки зору ефективності та масштабованості. Це означає можливість швидкого опрацювання великої кількості

письмових робіт без значного навантаження на викладачів. Водночас, з огляду на виявлену систематичну похибку в оцінюванні, доцільним залишається поєднання швидкості ШІ з аналітичною точністю людського контролю. Таким чином, отримані результати дозволяють зробити висновок про те, що штучний інтелект у поточному вигляді не є повністю еквівалентним людському оцінюванню. Він демонструє тенденцію до завищення балів, особливо у випадку слабких робіт, що може впливати на об'єктивність підсумкових результатів. Водночас його висока узгодженість із людьми у випадку якісних есе свідчить про потенціал використання таких систем як допоміжного інструменту.

Отримані дані підтверджують доцільність інтеграції Google Gemini в освітні практики не як заміни викладача, а як інструменту підтримки. Дана модель ефективно виконує функції первинного оцінювання, надає розгорнутий зворотний зв'язок і значно скорочує часові витрати, що є критично важливим у умовах великого навчального навантаження. Водночас остаточне оцінювання, особливо у випадках низької якості робіт або неоднозначних відповідей, потребує участі людини для забезпечення валідності та справедливості результатів.

Крім того, результати дослідження вказують на необхідність подальшого вдосконалення алгоритмів оцінювання, зокрема в аспектах чутливості до грубих помилок, відповідності темі та роботи з рукописним текстом. Перспективним напрямом є розробка адаптивних моделей, здатних калібрувати власні оцінки відповідно до людських стандартів, а також більш точно інтерпретувати контекст і якість письмового мовлення.

Отже, штучний інтелект уже сьогодні може виступати ефективним інструментом у системі оцінювання письмових робіт, однак його використання потребує методично обґрунтованого підходу, що поєднує автоматизацію з експертним людським контролем.

#### СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Олексенко С., Хоменко О. Реалізація потреби в професійному розвитку вчителів суспільствознавчих дисциплін в умовах кризи. *Der Stand der Entwicklung von Wissenschaft und Technik im XXI Jahrhundert* : monograph. Карлсруе, Німеччина: ScientificWorld-NetAkhatAV, 2023. С. 113–124.
2. Manning J., Baldwin J., Powell N. (2025) Human versus machine: The effectiveness of ChatGPT in automated essay scoring. *Innovations in Education and Teaching International*, 62(5), 1500–1513.
3. Tack A. (2025) Automated Essay Scoring. *Computational Linguistics*, 51(3), 1005–1008.
4. Hussein M. A., Hassan H., Nassef M. (2019) Automated language essay scoring systems: a literature review. *PeerJ Computer Science*, 5, e208.
5. Liu Y., Qi H., Lu X. (2025) Enhancing GPT-based automated essay scoring: the impact of fine-tuning and linguistic complexity measures. *Computer Assisted Language Learning*, 1–20.
6. García-Varela F., Nussbaum M., Mendoza M. (2025) ChatGPT as a Stable and Fair Tool for Automated Essay Scoring. *Education Sciences*, 15(8), 946.

7. Fontanilla K. P., Aquino J., Manzano C. et al. (2025) Effectiveness of Artificial Intelligence-Supported Academic Writing Tools with Peer Feedback in Enhancing ESL Students' Written Discourse: Basis in Designing Student-Oriented Training Program. *Student Journal of Technology and Educational Research*, 2(2), 109.
8. Zhou B., Pek L. S., Li J. et al. (2025) The Use of Artificial Intelligence Tools in English Academic Writing among University Students: A Scoping Review. *Language Teaching Research Quarterly*, 53, 95–114.
9. SapienPro (n.d.) AI Language Learning App Development. Available at: <https://sapien.pro/cases/ai-language-learning-app-development> (Accessed: 29 March 2026).
10. Das Deep P., Chen Y. (2025) The Role of AI in Academic Writing: Impacts on Writing Skills, Critical Thinking, and Integrity in Higher Education. *Societies*, 15(9), 247.

#### REFERENCES

1. Oleksenko S., Khomenko O. (2023) Realizatsiia potreby v profesiinomu rozvytku vchyteliv suspilstvoznavchykh dystsyplin v umovakh kryzy [Implementation of the need for professional development of social science teachers in crisis conditions]. *Der Stand der Entwicklung von Wissenschaft und Technik im XXI Jahrhundert: Monograph*, Karlsruhe, Germany: ScientificWorld-NetAkhatAV, 113–124. [in Ukrainian].
2. Manning J., Baldwin J., Powell N. (2025) Human versus machine: The effectiveness of ChatGPT in automated essay scoring. *Innovations in Education and Teaching International*, 62(5), 1500–1513.
3. Tack A. (2025) Automated Essay Scoring. *Computational Linguistics*, 51(3), 1005–1008.
4. Hussein M. A., Hassan H., Nassef M. (2019) Automated language essay scoring systems: a literature review. *PeerJ Computer Science*, 5, e208.
5. Liu Y., Qi H., Lu X. (2025) Enhancing GPT-based automated essay scoring: the impact of fine-tuning and linguistic complexity measures. *Computer Assisted Language Learning*, 1–20.
6. García-Varela F., Nussbaum M., Mendoza M. (2025) ChatGPT as a Stable and Fair Tool for Automated Essay Scoring. *Education Sciences*, 15(8), 946.
7. Fontanilla K. P., Aquino J., Manzano C. et al. (2025) Effectiveness of Artificial Intelligence-Supported Academic Writing Tools with Peer Feedback in Enhancing ESL Students' Written Discourse: Basis in Designing Student-Oriented Training Program. *Student Journal of Technology and Educational Research*, 2(2), 109.
8. Zhou B., Pek L. S., Li J. et al. (2025) The Use of Artificial Intelligence Tools in English Academic Writing among University Students: A Scoping Review. *Language Teaching Research Quarterly*, 53, 95–114.
9. SapienPro (n.d.) AI Language Learning App Development. Available at: <https://sapien.pro/cases/ai-language-learning-app-development> (Accessed: 29 March 2026).
10. Das Deep P., Chen Y. (2025) The Role of AI in Academic Writing: Impacts on Writing Skills, Critical Thinking, and Integrity in Higher Education. *Societies*, 15(9), 247.

Дата першого надходження статті до видання: 04.03.2026

Дата прийняття статті до друку після рецензування: 25.03.2026

Дата публікації (оприлюднення) статті: 19.05.2026

Стаття поширюється на умовах ліцензії відкритого доступу (CC BY 4.0)

