

О.В. Бахрушин, В.Є. Бахрушин

ЗАСТОСУВАННЯ ІНСТРУМЕНТІВ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ПОШУКУ ТА ПОПЕРЕДНЬОГО АНАЛІЗУ ДАНИХ У ПРИКЛАДНИХ ДОСЛІДЖЕННЯХ ОСВІТНІХ СИСТЕМ

Анотація. У статті окреслено проблеми застосування популярних інструментів штучного інтелекту GROK 3 та ChatCPT 4 для збирання наявних в інтернет даних про освітні системи, їх попередньої обробки й аналізу. Показано, що обидва інструменти дають змогу істотно прискорити і спростити виконання рутинних етапів досліджень, але GROK 3 дає змогу отримати більш повні, надійні і якісні дані, що може бути пов'язано як з можливістю доступу до актуальних даних мережі Інтернет, так і з відмінностями алгоритмів.

Ключові слова: штучний інтелект, система освіти, мережа закладів, фінансування, показники охоплення, збирання даних, описова статистика, надійність, релевантність, достовірність

Вступ. Постановка проблеми. Розроблення та застосування штучного інтелекту є одним з найбільш актуальних напрямів досліджень та розробок у сфері інформаційних технологій. Існує як багато ентузіазму, так і багато застережень стосовно застосування інструментів штучного інтелекту у дослідженнях та освітньому процесі. З одного боку, поширеною є думка, що штучний інтелект здатний самостійно розв'язувати будь-які прикладні, а можливо і наукові проблеми. З іншого боку, не менш поширеним є заперечення будь-яких спроб застосування інструментів штучного інтелекту не лише у науці, але і в освіті. Разом з тим, все більше фахівців схилиються до більш зваженого ставлення до штучного інтелекту і зосереджуються на вдосконаленні моделей та засобів штучного інтелекту, а також пошуку відповідей на питання: для яких задач, як саме і які саме засоби штучного інтелекту доцільно використовувати, якими є обмеження та можливі помилки при використанні конкретних інструментів для розв'язування конкретних типів задач, які проблеми забезпечення академічної доброчесності виникають при застосуванні штучного інтелекту тощо. Існує два основні напрями досліджень у сфері застосування методів штучного інтелекту у збиранні та аналізі даних про складні системи. Перший – це розроблення і вдосконалення моделей, алгоритмів, методів та інструментів для розв'язання відповідних задач. Другий – це застосування вже наявних інструментів для розв'язування прикладних задач збирання та аналізу даних щодо конкретних складних систем. У цій статті ми розглядаємо проблему другого типу щодо здатності поширених інструментів GROK 3 та ChatCPT 4

вирішувати подібні проблеми стосовно даних про вищу освіту України, країн ЄС і деяких інших країн.

Аналіз останніх досліджень і публікацій. У роботах [1,2] проаналізовано статистичні показники мережі закладів вищої освіти, фінансування, охоплення вищою освітою в Україні порівняно із країнами ЄС та ОЕСР за даними Державної служби статистики України та статистичних даних Євростат, ОЕСР, ЮНЕСКО та деяких інших джерел. Аналіз даних про мережу закладів вищої освіти України також було надано в [3,4]. Складні та суперечливі проблеми застосування штучного інтелекту у дослідженнях аналізуються у статті [5,6]. З одного боку, штучний інтелект розглядається як трансформаційна сила, яка може революціонізувати процеси проведення досліджень, мислення й написання академічних текстів, партнер в інтелектуальних пошуках, здатний усунути упередження та розширити людські здібності. З іншого боку – обговорюються питання меж впливу штучного інтелекту, пов'язані з його використанням етичні, епістемологічні та онтологічні проблеми. У роботі [7] зроблено огляд публікацій 2007-2018 р., які свідчать, що принаймні у сфері STEM освіти, інструменти штучного інтелекту можуть бути корисні для опрацювання наявної в різних джерелах інформації з метою персоналізації навчання, прогнозування його результатів, більш об'єктивного оцінювання тощо. Автори огляду [8] на підставі аналізу великої кількості інших досліджень останніх років роблять висновок, що використання штучного інтелекту може збільшити знання про процеси, які проявляються в освіті на національному, регіональному та локальному рівнях, допомагаючи сформулювати політику та розподілити ресурси для характеристик, визначених як важливі для ефективності освіти (макроперспектива). Водночас, можливі більш конкретні дослідження окремих учнів, вчителів і шкіл (мікроперспектива), орієнтовані на персоналізацію навчання, вдосконалення педагогічних стратегій, прогнозування короткострокових наслідків тих чи інших дій тощо. Значна частина публікацій присвячена впливу штучного інтелекту взагалі чи конкретних інструментів на успішність та результати навчання на різних рівнях освіти. Водночас, у прикладних дослідженнях недостатньо уваги приділено оцінюванню достовірності, надійності та релевантності результатів, отриманих за допомогою штучного інтелекту, зокрема таких інструментів, як GROK 3 та ChatCPT 4.

Мета дослідження: формулювання окремих рекомендацій щодо застосування інструментів штучного інтелекту GROK 3 та ChatCPT 4 для розв'язування деяких типових задач збирання та аналізу даних великого обсягу і складної структури на прикладі даних про системи вищої освіти різних країн.

Викладення основного матеріалу дослідження. Для проведення дослідження було обрано декілька актуальних задач, що можуть розглядатися як типові при проведенні прикладних досліджень освітніх систем, а саме:

- збирання даних щодо кількості студентів у закладах вищої освіти країн ЄС, Австралії, Канади, США і Японії та оцінювання основних статистичних характеристик відповідних розподілів;

– збирання даних щодо фінансування здобуття вищої освіти у розрахунку на студента у закладах вищої освіти країн ЄС, Австралії, Канади, США і Японії та оцінювання основних статистичних характеристик відповідних розподілів;

– збирання даних щодо показників залучення громадян до здобуття вищої освіти.

Варто зазначити, що зазначені задачі не є змістовно складними для фахівців. Основні складності полягають у тому, що не існує єдиних, актуальних та стандартизованих наборів відповідних даних. Наявні дані розпорошені по різних джерелах та можуть стосуватися різних років і мати особливості, що відображають специфічні особливості різних країн, зумовлені методиками збирання первинних даних, відмінностями термінології тощо. GROK 3 та ChatGPT 4 використовують істотно різні підходи до розв’язування таких задач. Як приклад розглянемо відповіді на Запитання 1: «Please give me a table containing such data (by country) for USA, Canada, Japan, Australia and each of EU countries: number of higher education institutions, mean and median values of number of students in institutions of relative country, standard deviations for these distributions, and minimum/maximum values of number of students».

Як GROK 3 так і ChatGPT 4 надають відповідь протягом кількох секунд, але відповідь GROK 3 є значно більш повною та змістовною. Вона також викликає більшу довіру, зокрема і через те, що краще описана методика, використано більш актуальні дані, а наведені у відповіді результати збігаються з результатами вибіркової ручної перевірки. Крім того, GROK 3 додає посилання на сайти або сторінки, звідки було взято вихідну інформацію.

У відповіді на це запитання GROK 3 наводить таблицю запрошених даних, а також додаткову інформацію по методиках та джерелах даних. На рис. 1 наведено фрагмент таблиці з відповіді GROK 3.

Table: Higher Education Institutions Data (Approximate)						
Country	# of Institutions	Mean Students	Median Students	Std. Deviation	Min Students	Max Students
USA	3,228 (2019)	9,000	5,000	15,000	100	200,000
Canada	223 (2023)	12,000	8,000	10,000	200	80,000
Japan	784 (2023)	6,000	4,000	8,000	150	50,000
Australia	172 (2023)	15,000	10,000	12,000	300	70,000
EU Countries						
Austria	70 (2022)	5,000	3,500	6,000	100	40,000
Belgium	45 (2022)	8,000	6,000	7,000	150	50,000

Рисунок 1 - Фрагмент відповіді GROK 3 на запитання 1

Зокрема, зазначено, що під вищою освітою розуміються ступеневі програми, які відповідають 5-8 рівням Міжнародної стандартної класифікації освіти. Згідно з відповіддю, дані про кількість закладів отримано з відкритих баз Інституту статистики

ЮНЕСКО та ОЕСР, а також з даних національної статистики. З них використано останні доступні на дату запиту (08.04.2025) дані, які переважно стосуються 2023-2024 років. Середні значення отримано як відношення загальної кількості студентів до кількості ЗВО. Медіанні значення (середнє значення), де це можливо, оцінені за доступними розподілами або з урахуванням типовій для освітніх систем асиметрії розподілу. Стандартні відхилення у більшості випадків недоступні і оцінювалися на основі загальних тенденцій варіабельності, приміром, співвідношення кількості великих і малих закладів. Так само оцінювали мінімальний та максимальний обсяги, якщо точні дані не були доступні. GROK 3 також надає додаткову інформацію по країнах. Приміром, зазначено, що для США розподіл спотворюють деякі великі університети, як університет Аризони; дані по кількості університетів стосуються 2019 р., а дані по кількості студентів – 2023 р. У відповіді зазначено особливості, що могли вплинути на точність оцінок та на зіставність результатів по окремих країнах. Це стосується неповноти даних по окремих країнах, статичності даних у джерелах інформації, різних визначень закладу вищої освіти.

Слід зазначити, що GROK 3 намагається мінімізувати обсяг своєї роботи, якщо це дозволяє формулювання завдання. Зокрема, при заміні у завданні «each of EU countries» на «EU countries» у відповіді було наведено результати лише по декількох країнах ЄС. Дані, що є у інтегрованих статистичних базах GROK 3 бере з них. При цьому інколи він користується не первинними базами Євростат, ОЕСР тощо, а вторинними джерелами, що беруть інформацію з цих баз. Зазвичай це не створює проблем.

Відповідь ChatGPT 4 на те саме запитання є досить лаконічною. В ній також, як і у відповіді GROK 3, повідомляється про неповноту даних по окремих країнах, відмінності в методах збирання даних, звітних періодах і визначеннях ЗВО. На рис. 2 наведено фрагмент відповіді, яку ChatGPT 4 навів у текстовому, а не у табличному (тобто не відповідно до запиту) форматі.

Аналогічні відповіді ChatGPT 4 надав по США і Канаді. При цьому варто зазначити, що для цих країн було наведено не розраховані за актуальними даними середні значення кількості студентів у ЗВО, а наявні у базі історичні дані кінця 1980х – початку 1990-х років безпосередньо про середні значення.

Запитані результати по країнах ЄС Chat GPT 4 не надав, посилаючись на недоступність вичерпних та уніфікованих даних. Натомість він порадив проконсультуватися з відповідними національними міністерствами освіти або статистичними агентствами кожної країни ЄС. Судячи з відповідей, Chat GPT 4 спочатку намагався знайти у своїй базі готові дані про запитувані показники для будь-якого року і лише за їх відсутності намагався додатково виконати найпростіші розрахунки (середні арифметичні для Японії та Австралії). Більш складні розрахунки він не намагався виконувати.

Аналогічною є ситуація при відповіді на запитання «Please give me following data (by country) in tabular format for Ukraine, the USA, Canada, Japan, Australia and each EU country in 1990, 2000, 2010, 2020: gross enrolment ratio in tertiary education; share of citizens aged 25-34 with a bachelor's degree or higher». Це завдання є більш простим, оскільки переважна більшість запитаних даних у розрізі зазначених країн та років є у великих

статистичних базах Євростат, ОЕСР, Інституту статистики ЮНЕСКО та Світового банку.

GROK 3 надає таблицю запитаних даних, що містить окремі пропуски (12,9% від загальної кількості, з яких майже всі стосуються 1990 р.). Наведені дані є близькими до даних інших джерел. Також майже повною є відповідь на додаткове запитання про відсотки громадян, що мають ступінь бакалавра і ступінь магістра у віковій групі 25-34 роки. Як і у попередньому випадку GROK 3 наводить джерела інформації (переважно зазначені вище великі бази), окремі коментарі до результатів і додаткові дані. Зокрема, вказано, що агреговані дані по ЄС, які надані додатково до даних по країнах, для 1990 та 2000 р. не враховують відсутні дані по окремих країнах, що окремі спеціально позначені дані є оцінками, отриманими екстраполяцією, що для Японії замість неповних даних 2020 р. наведено дані за 2019 р., тощо.

Japan:

- **Number of Higher Education Institutions:** As of May 1, 2020, Japan had 795 universities. [citeturn0search4](#)
- **Total Student Enrollment:** In 2023, approximately 2.95 million students were enrolled in universities. [citeturn0search5](#)
- **Average Number of Students per Institution:** Based on 2023 data, the average is approximately 3,710 students per university.

Australia:

- **Number of Higher Education Institutions:** As of recent reports, Australia has 43 universities, comprising 37 public and 6 private institutions. [citeturn0search19](#)
- **Total Student Enrollment:** In 2021, there were 1,185,450 students attending university or other higher education institutions. [citeturn0search16](#)
- **Average Number of Students per Institution:** Approximately 27,570 students per university, based on 2021 data.

European Union (EU) Countries: Comprehensive and uniform data across all EU member states are not readily available due to differing data collection standards and reporting

Рисунок 2 - Фрагмент відповіді ChatGPT 4 на запитання 1

Натомість Chat GPT 4 взагалі не перевіряє зазначені великі бази, що може бути зумовлено відсутністю доступу до актуальної інформації через Інтернет і необхідністю обсягу оптимізації інформації, що зберігається у локальній базі. Тож не дивно, що єдиними результатами, які були видані, є оцінки GER для України для 2010 та 2020 р., які є дещо відмінними, але близькими до відповідних даних GROK 3. У всіх інших випадках було зазначено, що дані відсутні.

Аналогічною є ситуація із відповіддю на запитання: «Please give me a table containing such data (by country) for USA, Canada, Japan, Australia and each EU countries : number of higher education institutions, mean and median values (in nominal USD per stu-

dent) of general, public and private funding in institutions of relative country, standard deviations for these distributions, and minimum/maximum values of such funding». GROK 3 надає повну інформацію, яка у частині середніх по країнах даних є близькою до даних інших джерел. Chat GPT 4 надає суперечливі відповіді на це запитання.

Висновки:

1. Результати дослідження свідчать про те, що GROK 3 та ChatGPT 4 дають змогу істотно прискорити і спростити виконання рутинних етапів збирання та попереднього аналізу даних великого обсягу та складної структури при дослідженні освітніх систем.

2. Характерні для освітніх систем неповнота даних по різних країнах та роках, різні структури та методики збирання/розрахунку подібних даних, відмінності термінології тощо створюють ризики недостовірності, ненадійності і нерелевантності результатів роботи штучного інтелекту. Для GROK 3 порівняно з ChatGPT 4 ці ризики є значно меншими, оскільки він без додаткових запитань наводить у відповіді інформацію про методологічні проблеми і пов'язані з ними застереження. Втім, навіть і в цьому випадку варто робити верифікацію його даних їх зіставленням з відомими або отриманими в інший спосіб даними.

3. Деякі проблеми ChatGPT 4, очевидно, пов'язані з тим, що він не має доступу до актуальних даних мережі Інтернет і працює з даними локальної бази. Такі дані необхідно оптимізовувати з метою забезпечення можливості надання «хороших» відповідей на запитання, які за прогнозами будуть найбільш масовими. Але через це знижується якість виконання вузькоспеціальних завдань. Зокрема, інколи ChatGPT 4 витрачає час на пошук інформації в нерелевантних базах, а в інших випадках не знаходить інформацію, наявну у таких популярних джерелах, як бази Євростат, ОЕСР, Інституту статистики ЮНЕСКО, Світового банку.

4. GROK 3 містить більше інструментів попередньої обробки даних. Тому він самостійно розраховує або оцінює окремі показники описової статистики за наявності навіть неповних вибірок. ChatGPT 4 зазвичай бере лише наявні в джерелах готові дані, навіть, якщо вони дуже застарілі. Єдиний вид розрахунків, який він в окремих випадках виконував у розглянутих прикладах – це визначення середнього арифметичного діленням сумарного (по країні) значення показника на кількість закладів.

ЛІТЕРАТУРА

1. Бахрушин В.Є. Модернізація мережі закладів вищої освіти України: Деякі євроінтеграційні аспекти. Освітня аналітика України, 2023, № 5, с. 112-125. URL: <https://doi.org/10.32987/2617-8532-2023-5-112-125>
2. Бахрушин В.Є. Особливості фінансування вищої освіти в Україні порівняно з іншими країнами. В кн. «Удосконалення фінансування системи освіти України як передумова успішного реформування: монографія / за ред. С. Л. Лондара; ДНУ «Інститут освітньої аналітики». Київ, 2021, с. 251-264.
URL: https://iea.gov.ua/wp-content/uploads/2024/05/monograph_20212_fin.pdf
3. Кремень В. Г., Луговий В. І., Саух П. Ю., Таланова Ж. В. Мережа державних закладів вищої освіти України: аналітичний огляд конкурентоспроможності. Вісник

Національної академії педагогічних наук України, 2022, Т. 4, № 1, С. 1-8.
URL: <https://doi.org/10.37472/v.naes.2022.4122>

4. Калашнікова С., Оржель О. Оптимізація мережі закладів вищої освіти: теоретичні особливості та практичні рекомендації. *Університети і лідерство*. 2022. № 13. С. 89–129. URL: <https://doi.org/10.31874/2520-6702-2022-13-89-129>

5. Butson, R., & Spronken-Smith, R. AI and its implications for research in higher education: a critical dialogue. *Higher Education Research & Development*, 2024, 43(3), 563–577. <https://doi.org/10.1080/07294360.2023.2280200>

6. Bond, M., Khosravi, H., De Laat, M. et al. A meta systematic review of artificial intelligence in higher education: a call for increased ethics, collaboration, and rigour. *Int. J. Educ. Technol. High Educ.*, 2024, 21:4, 41 p. <https://doi.org/10.1186/s41239-023-00436-z>

7. Zawacki-Richter, O., Marín, V.I., Bond, M. et al. Systematic review of research on artificial intelligence applications in higher education – where are the educators? *Int. J. Educ. Technol. High Educ.*, 2019, 16:39, 27 p. <https://doi.org/10.1186/s41239-019-0171-0>

8. de Souza Zanirato Maia, J.; Bueno, A.P.A.; Sato, J.R. Applications of Artificial Intelligence Models in Educational Analytics and Decision Making: A Systematic Review. *World*, 2023, 4, 288-313. <https://doi.org/10.3390/world4020019>

REFERENCES

1. Bakhrushyn V. Modernization of the network of higher education institutions in Ukraine: some aspects of European integration, 2023, № 5, p. 112-125. <https://doi.org/10.32987/2617-8532-2023-5-112-125>

2. Bakhrushyn V. Features of financing higher education in Ukraine compared to other countries. In «Improving the financing of the education system of Ukraine as a prerequisite for successful reform» : monograph / Ed. S.D. Londar; State Scientific Institution “Institute of Educational Analytics”. Kyiv, 2021, p. 251-264.

URL: https://iea.gov.ua/wp-content/uploads/2024/05/monograph_20212_fin.pdf

3. Kremen V., Lugovyi V., Saukh P., Talanova Z. Network of state higher education institutions in Ukraine: an analytical review of competitiveness. *Herald of the National Academy of Educational Sciences of Ukraine*, 2022, Vol. 4, № 1, p. 1-8.
<https://doi.org/10.37472/v.naes.2022.4122>

4. Kalashnikova S., Orzhel O. Optimization of the Network of Higher Education Institutions: Theoretical Features and Practical Recommendations. *International Scientific Journal of Universities and Leadership*. 2022. № 13. p. 89–129. DOI: <https://doi.org/10.31874/2520-6702-2022-13-89-129>

5. Butson, R., & Spronken-Smith, R. AI and its implications for research in higher education: a critical dialogue. *Higher Education Research & Development*, 2024, 43(3), 563–577. <https://doi.org/10.1080/07294360.2023.2280200>

6. Bond, M., Khosravi, H., De Laat, M. et al. A meta systematic review of artificial intelligence in higher education: a call for increased ethics, collaboration, and rigour. *Int. J. Educ. Technol. High Educ.*, 2024, 21:4, 41 p. <https://doi.org/10.1186/s41239-023-00436-z>

7. Zawacki-Richter, O., Marín, V.I., Bond, M. et al. Systematic review of research on artificial intelligence applications in higher education – where are the educators? *Int. J. Educ. Technol. High Educ.*, 2019, 16:39, 27 p. <https://doi.org/10.1186/s41239-019-0171-0>

8. de Souza Zanirato Maia, J.; Bueno, A.P.A.; Sato, J.R. Applications of Artificial Intelligence Models in Educational Analytics and Decision Making: A Systematic Review. World, 2023, 4, 288-313. <https://doi.org/10.3390/world4020019>

Received 15.05.2025.
Accepted 21.05.2025.

Application of artificial intelligence tools for data search and preliminary analysis in applied research of education systems

The article outlines the problems of using popular artificial intelligence tools GROK 3 and ChatCPT 4 for collecting data on educational systems available on the Internet, their pre-processing and analysis. A significant proportion of recent publications is devoted to the development of new artificial intelligence tools, the study of algorithms of its work, methodological and ethical problems of its application in various fields, the impact of artificial intelligence in general or of specific tools on academic success and learning outcomes at different levels of education. At the same time, applied research has not paid enough attention to assessing the validity, reliability, and relevance of results obtained using artificial intelligence, in particular such tools as GROK 3 and ChatCPT 4.

For this study, several problems of educational systems researches, which can be considered as relevant also in broader contexts, were selected. The main attention was paid to the issues of collecting and preliminary statistical analysis of data on the number of students in higher education institutions, funding of higher education, and indicators of youth involvement in higher education.

It has been shown that both tools significantly speed up and simplify the implementation of routine research stages, but GROK 3 allows to obtain more complete, reliable and high-quality data, which may be due to both the ability to access current data on the Internet and differences in algorithms. ChatCPT 4 is usually limited to providing a small amount of the requested data present in its database, which is sometimes very outdated. The advantage of GROK 3 is also that it collects not only the requested numerical data from various types of sources, but is also able to independently calculate simple statistical characteristics of samples, such as mean values, medians, standard deviations, minimum and maximum values, from the primary data without additional questions, including the ability to make plausible estimates from incomplete, highly asymmetric samples. Nevertheless, in this case too, the general methodology should provide for verification of the results of artificial intelligence, in particular by comparing them with other known data.

Бахрушин Олексій Володимирович – аспірант кафедри програмних засобів Національного університету «Запорізька політехніка».

Бахрушин Володимир Євгенович – д.ф.-м.н., професор кафедри системного аналізу та обчислювальної математики Національного університету «Запорізька політехніка».

Bakhrushyn Oleksiy Volodymyrovych – postgraduate student of the Department of Software Tools of the National University “Zaporizhzhia Polytechnic”.

Bakhrushyn Volodymyr Yevgenovych – Doctor of Physical and Mathematical Sciences, Professor of the Department of System Analysis and Computational Mathematics of the National University “Zaporizhzhia Polytechnic”.