

ТЕНДЕНЦІЇ РОЗВИТКУ ДОСЛІДЖЕНЬ У СФЕРІ ШТУЧНОГО ІНТЕЛЕКТУ В ОСВІТІ: ТЕМАТИЧНЕ МОДЕЛЮВАННЯ ЗА ДОПОМОГОЮ ЛАТЕНТНОГО РОЗПОДІЛУ ДІРІХЛЕ

Філь Н.Ю.¹, Слісаренко Р.В.², Дейнеко Ж.В.², Морозова Л.Ю.²

¹Харківський національний автомобільно-дорожній університет

²Харківський національний університет радіоелектроніки

Анотація. Штучний інтелект ініціює радикальну трансформацію освітнього середовища, модифікуючи методології навчання, механізми оцінювання та інструменти персоналізації освітніх процесів. У цьому дослідженні проаналізовано наукові тренди інтеграції штучного інтелекту в освіту з використанням тематичного моделювання. На підставі даних із Web of Science виокремлено перспективи розвитку інтелектуальних освітніх технологій, що становлять основу для формування інноваційних освітніх парадигм.

Ключові слова: штучний інтелект, освіта, персоналізація, машинне навчання, латентний розподіл Діріхле, тематичне моделювання.

Вступ

Сучасний розвиток штучного інтелекту (ШІ) трансформує сферу освіти, пропонуючи персоналізацію навчання, індивідуальні плани для студентів, імерсивне навчання та інтелектуальне відстеження прогресу. Технології адаптивного навчання знижують навантаження на викладачів, даючи їм зосередитись на роботі із студентами [1]. Однак упровадження ШІ в освіту породжує питання щодо ефективності, методологічних підходів до дослідження його впливу й глобальних тенденцій наукової співпраці.

Незважаючи на зростання інтересу до AIED (Artificial Intelligence in Education) – міждисциплінарної галузі, що поєднує комп'ютерні науки, освіту та когнітивну психологію [2], спостерігається фрагментарність підходів до вивчення цього феномену. Дослідження часто фрагментовані, зосереджені на вузьких аспектах, що ускладнює формування цілісної уяви про тенденції та перспективи розвитку AIED. Відсутність узагальненої картини міжнародної наукової взаємодії ускладнює оцінювання впливу міждисциплінарних досліджень на освітні концепції.

Актуальним є комплексний аналіз досліджень з використанням методів наукової аналітики, таких як тематичне моделювання та мережевий аналіз. Це дасть змогу не лише окреслити ключові завдання досліджень у цій сфері, а й визначити перспективні напрями розвитку AIED, що можуть стати основою для формування нових освітніх парадигм у контексті ШІ.

Аналіз публікацій

У сучасних дослідженнях, присвячених використанню ШІ в освіті, розглядаються як можливості, так і виклики впровадження зазначеної технології. Зокрема увагу зосереджено на персоналізації навчання, автоматизації оцінювання, питаннях етики та цифрової нерівності.

Одним із ґрунтовних досліджень є робота Wolfgang Nejdl, Christoph Rensing, Santi Caballé, Josep Lluís Larriba-Pey, де проаналізовано персоналізацію навчального процесу за допомогою ШІ. Автори доводять, що адаптивні алгоритми можуть покращити ефективність навчання, автоматично налаштовуючи контент відповідно до рівня знань студентів [3].

Проте автори звертають увагу на ризики алгоритмічних упереджень і необхідність прозорості в роботі ШІ-систем.

Giuseppe Rizzo, Giuseppe De Pietro, Luigi De Pietro розглядають можливості автоматизації оцінювання та адаптивного навчання. Вони вважають необхідними регуляторні механізми для захисту персональних даних студентів.

Автори також наголошують, що без належного регулювання ШІ може стати інструментом маніпуляції або дискримінації в навчальному процесі [4].

Pierre Dillenbourg, Mikhail M. Kasyanov присвятили роботу перспективам індивідуалізації навчання завдяки ШІ. Автори стверджують, що адаптивні технології здатні значно покращити засвоєння матеріалу, проте їх використання не має призводити до заміни

викладачів. Додатково вони звертають увагу на проблему цифрової нерівності, що може обмежувати доступ до високоякісних ШІ-рішень для студентів із соціально вразливих груп [5].

Wayne Holmes, Maya Bialik, Charles Fadel звертають увагу на етичних аспектах і нормативному регулюванні впровадження ШІ в освітній процес.

Вони наголошують на необхідності навчання педагогів ефективного використання технологій, а також на розробленні стандартів прозорості та відповідальності в процесі впровадження ШІ в освітні установи [6].

Попри вагомій напрацюванні у сфері застосування ШІ в освіті, залишаються нерозв'язаними деякі питання, зокрема:

- відсутність комплексного аналізу динаміки досліджень у цій сфері на основі великомасштабних даних;
- недостатність систематичних підходів до виявлення трендів у розвитку AIED;
- необхідність розроблення універсальної методології оцінювання ефективності впровадження ШІ в освіті.

Ці питання є ключовими для нашого дослідження. У межах цієї роботи буде здійснено збір і попереднє оброблення даних із *Web of Science*, використано модель для аналізу трендів, інтерпретовано досягнуті результати та візуалізовано їх.

Це дасть змогу виявити основні напрями розвитку AIED, оцінити популярність різних підходів та визначити перспективи їх подальшого використання.

Мета й постановка завдання

Метою дослідження є проведення комплексного аналізу сучасних наукових досліджень у сфері (AIED) за допомогою збору, систематизації та оброблення інформації з міжнародних наукових баз даних, зокрема *Web of Science*, для виявлення ключових наукових трендів, оцінювання динаміки розвитку AIED та визначення перспективних напрямів подальших досліджень.

Для досягнення поставленої мети необхідно систематизувати та проаналізувати наукові публікації з тематики AIED, визначити основні тренди та напрями досліджень у цій сфері, оцінити динаміку розвитку напрацювань за останні роки, розробити підходи до візуалізації досягнутих результатів для полегшення їх інтерпретації та аналізу, а також запропонувати рекомендації щодо використання результатів дослідження

в освітній політиці та науковій діяльності.

Методологія тематичного моделювання в аналізі AIED: порівняння LDA та NMF

Для аналізу сучасних досліджень у сфері AIED було вирішено впровадити методи тематичного моделювання, зокрема *Latent Dirichlet Allocation* (LDA) та *Non-negative Matrix Factorization* (NMF). Обидва методи є потужними інструментами для виявлення латентних структур у текстових даних, що дають змогу структурувати корпуси наукових публікацій та ідентифікувати ключові напрями досліджень. Проте вони ґрунтуються на різних математичних підходах, що робить їх придатними для будь-яких аналітичних задач.

LDA – це генеративна ймовірнісна модель корпусу. Основна ідея полягає в тому, що документи подані як випадкові суміші латентних тем, де кожна тема визначається розподілом слів [7]. Ймовірнісна модель LDA припускає, що кожен документ d з D документів у корпусі може бути описаний як ймовірнісна суміш T тем. Ці ймовірності містяться у векторі тем документа (θ_d) – завдовжки T . Результатом LDA є матриця θ розміром $D \times T$, яка містить $P(t_1|d_d)$ – ймовірність того, що документ d належить до теми t :

$$\theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_D \end{pmatrix} = \begin{pmatrix} P(t_1|d_1) & \cdots & P(t_T|d_1) \\ \vdots & \ddots & \vdots \\ P(t_1|d_D) & \cdots & P(t_T|d_D) \end{pmatrix}, \quad (1)$$

де θ – матриця розподілів тем документів розміром $D \times T$; θ_i – вектор розподілу тем для i -го документа, де $i = 1, 2, \dots, D$; $P(t_j | d_i)$ – ймовірність того, що документ d_i належить до теми t_j ; D – загальна кількість документів у корпусі; T – загальна кількість тем; t_j – j -та тема, де $j = 1, 2, \dots, T$; d_i – i -й документ, де $i = 1, 2, \dots, D$.

Кожна тема $t \in T$ визначається ймовірнісним розподілом над словниковим запасом розміром (V) [8].

Отже, тема визначає ймовірність появи певної лексичної одиниці за умови належності до цієї теми. Наприклад, у документах, присвячених фізиці, спостерігається підвищена ймовірність використання термінів, зокрема «частинка», «квантовий» та «бозон», тоді як у медичних текстах переважають лексеми, наприклад «клінічний», «вірус» та «інфекція». Вектори ймовірностей слів

для кожної теми містяться в матриці (ϕ) розміром ($V \times T$):

$$\phi = (\phi_1, \dots, \phi_T) = \begin{pmatrix} P(w_1|t_1) & \dots & P(w_1|t_T) \\ \vdots & \ddots & \vdots \\ P(w_V|t_1) & \dots & P(w_V|t_T) \end{pmatrix}, \quad (2)$$

де ϕ – матриця розподілів слів за темами розміром $M \times T$; ϕ_T – вектор розподілу слів для теми t , де $t = 1, 2, \dots, T$; $P(w_v|t_t)$ – умовна ймовірність того, що слово w_v з'являється в темі t_t ; V – розмір словникового запасу (кількість унікальних слів у корпусі); T – загальна кількість тем; w_v – v -те слово зі словникового запасу, де $v = 1, 2, \dots, V$; t_t – t -та тема, де $t = 1, 2, \dots, T$.

Імовірності $P(w_v|t_t)$, що містяться у векторах ϕ_t , визначають умовну ймовірність появи лексичної одиниці v зі словникового корпусу в разі належності до теми t . Вектори ϕ_t дають змогу інтерпретувати семантичний зміст кожної теми, оскільки модель LDA не передбачає явного маркування тем.

Однак недоліками LDA є складність налаштування гіперпараметрів, а також висока обчислювальна складність для великомасштабних даних.

На відміну від LDA, NMF є декомпозиційним, неймовірнісним алгоритмом, що оснований на методах матричної факторизації та належить до класу лінійно-алгебраїчних підходів [9]. NMF застосовується до даних, попередньо трансформованих за допомогою TF-IDF, і полягає в розкладі вихідної матриці на дві матриці нижчого рангу [10]. TF-IDF зі свого боку є статистичною мірою, що оцінює значущість терміна в межах корпусу документів. Як ілюструє рис. 1, NMF здійснює декомпозицію термін-документної матриці (A) на добуток двох

матриць: матриці термінів-тем (W) та матриці тем-документів (H). Елементи матриць W та H оновлюються ітеративно, де W містить базисні вектори, а H – коефіцієнти їх лінійної комбінації [11]. Обмеження невід'ємності елементів W та H є суттєвим, оскільки наявність від'ємних значень ускладнює інтерпретацію семантичного змісту тем [12].

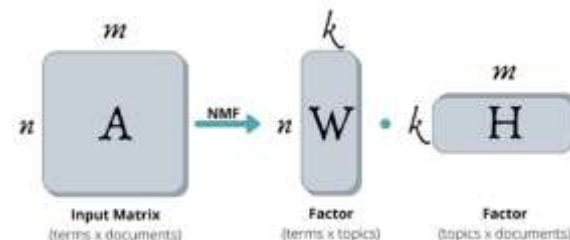


Рис. 1. Подання NMF

NMF менш гнучкий за умови семантичного перекриття тем, вимагає якісного попереднього оброблення інформації. Переваги та недоліки методів порівняно в табл. 1. Вибір між LDA та NMF залежить від специфіки даних та цілей дослідження. У нашому випадку, зважаючи на складність і багатогранність досліджень у сфері AIED, де теми часто перетинаються, LDA виглядає більш привабливим варіантом.

Аналіз результатів тематичного моделювання LDA

Для аналізу наукових публікацій у сфері AIED використано тематичне моделювання. Вхідні тексти було піддано глибокому лінгвістичному обробленню, що передбачало вилучення пунктуаційних символів, чисел та стоп-слів, а також нормалізацію лексем за допомогою лематизації.

Таблиця 1 – Порівняння основних переваг та недоліків LDA та NMF

Критерій	LDA	NMF
Інтерпретація результатів	Теми як імовірнісні розподіли слів, документи описуються як суміш тем.	Теми як лінійні комбінації слів з невід'ємними вагами.
Оброблення розріджених даних	Добре обробляє розріджені дані за допомогою ймовірнісного підходу.	Чутливий до розрідженості, вимагає якісного попереднього оброблення.
Масштабованість	Вимагає більше ресурсів для великих даних.	Швидше та ефективніше для великих даних.
Гнучкість	Краще підходить для текстів із перекриттям тем.	Краще для текстів із чіткими, неперекриваючими темами.
Легкість реалізації	Вимагає налаштування гіперпараметрів (α, β).	Проста реалізація з мінімальними налаштуваннями.
Інтерпретація результатів	Теми як імовірнісні розподіли слів, документи описуються як суміш тем.	Теми як лінійні комбінації слів з невід'ємними вагами.

Це забезпечило фільтрацію неінформативних елементів тексту та збереження ключових семантично значущих термінів, таких як іменники, дієслова, прикметники та прислівники, що є критично важливим для точного тематичного моделювання.

Далі корпус документів було перетворено в набір слів, що дало змогу подати кожен текст у вигляді розрідженого векторного подання та побудувати словник унікальних термінів.

Для навчання LDA-моделі було визначено п'ять тематичних кластерів, що узгоджується з емпірично обґрунтованим припущенням про наявність ключових напрямів дослідження в межах AIED.

Оптимізація параметрів здійснювалася способом багаторазового проходження корпусу текстів (35 ітерацій) із застосуванням баєсівського оцінювання латентних розподілів тем у документах.

Використання фіксованого випадкового стану (*random_state=42*) забезпечило відтворюваність результатів та стабільність тематичних розподілів, що критично важливо для коректного порівняння отриманих кластерів.

Запропонована модель допомагає не лише класифікувати документи за темами, а й оцінити вагові коефіцієнти ключових слів у кожній темі, що значно покращує інтерпретованість результатів у контексті освітніх досліджень.

Візуалізація та оцінювання міжтематичних зв'язків

На основі *Intertopic Distance Map* (див. рис. 2) можна зробити висновок, що отримані теми мають різний рівень перекриття.

Виокремлено кілька чітко розділених тематичних кластерів, що підтверджує гіпотезу про різноманітність досліджень у сфері AIED. Водночас деякі теми демонструють частковий перетин, що свідчить про міждисциплінарні особливості наукових досліджень.

Аналіз топ-30 найбільш значущих термінів (праворуч на першому зображенні) вказує на високу частоту вживання слів, пов'язаних із навчанням (*student*, *teacher*, *learning*, *course*), технологіями штучного інтелекту (*chatbot*, *algorithm*, *model*, *neural network*), а також медичними аспектами (*medical*, *clinical*, *patient*). Це свідчить про активний розвиток AIED у різних доменах: освіті, медицині, технологіях оцінювання знань.

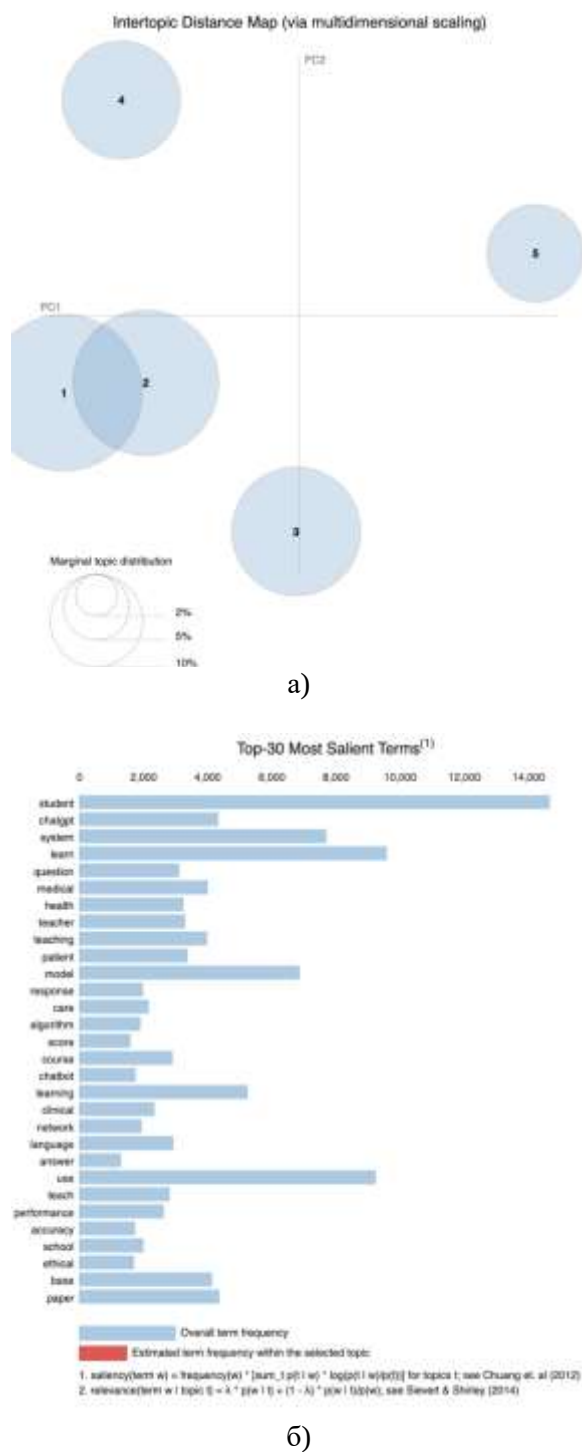


Рис. 2. Міжтематична карта відстаней: а – карта міжтематичних відстаней (за допомогою багатовимірнього масштабування); б – топ-30 найважливіших термінів у дослідженні

Розподіл документів у латентному просторі

Аналіз тематичних зв'язків у корпусі документів було виконано за допомогою t-SNE (*t-Distributed Stochastic Neighbor Embedding*), що дає змогу візуалізувати латентні структу-

ри у високовимірних даних (рис. 3). У цій візуалізації кожна точка відповідає окремому документу, а її колір позначає домінуючу тему.

Отримана карта демонструє чітке розділення тематичних кластерів, що вказує на добре визначені межі між певними темами. Найбільш концентровані регіони (наприклад, у правому нижньому куті) можуть свідчити про групу документів, які мають сильну тематичну схожість, що може бути зумовлено використанням специфічної термінології або дослідженням окремих вузькоспеціалізованих аспектів AIED. Водночас у центральній частині спостерігається плавний перехід між темами, що може бути свідченням міждисциплінарної особливості певних досліджень.

Особливо цікавим є взаємозв'язок між темами, що проявляється у вигляді зон змішаних кольорів і свідчить про можливе тематичне перекриття. Це властиво для міжгалузевих досліджень, де ШІ використовується в освіті, медицині, алгоритмічному моделюванні та соціальних науках. Таким чином, t-SNE підтверджує досягнуті раніше результати багатовимірного масштабування (MDS), демонструючи як чітко відокремлені теми, так і їх перехідні зони.



Рис. 3. t-SNE-візуалізація тем

Інтерпретація основних тем (Word Cloud)

Для аналізу отриманих тем використано хмару слів для кожної з виділених:

– «Штучний інтелект у медичній освіті» (рис. 4) об'єднує дослідження, зосереджені на ролі AIED у сфері медичної підготовки та охорони здоров'я. Найбільш значущі терміни, такі як *health*, *medical*, *patient*, *care*, *education*, *study*, *clinical*, *ethical*, вказують на активний розвиток автоматизованих навчальних платформ для медичних працівників.



Рис. 4. Тема 1: «Штучний інтелект у медичній освіті»

– «Штучний інтелект в освіті» (рис. 5) містить терміни *education*, *technology*, *research*, *artificial intelligence*, *development*, що свідчить про дослідження в напрямі інтеграції ШІ в навчальні процеси.



Рис. 5. Тема 2: «Штучний інтелект в освіті»

– «Алгоритмічні методи та машинне навчання в освіті» (рис. 6) охоплює дослідження алгоритмічних підходів, що використовуються в навчальних середовищах. Ключові слова *system*, *model*, *use*, *base*, *datum*, *method*, *algorithm*, *learning*, *intelligence*, *network* вказують на широкий спектр методів, зокрема нейронні мережі та оброблення освітньої інформації.

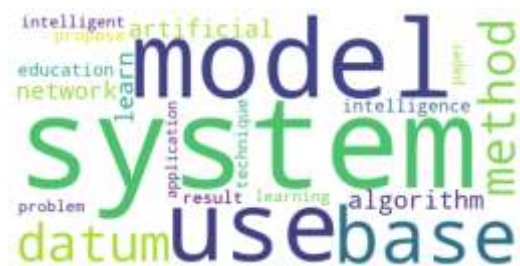


Рис. 6. Тема 3: «Алгоритмічні методи та машинне навчання в освіті»

– «Роль великих мовних моделей у навчанні» (рис. 7) охоплює дослідження щодо використання генеративного штучного інтелекту та чат-ботів у навчальних процесах. Серед основних термінів вирізняються *chatgpt*, *question*, *response*, *use*, *study*, *chatbot*,

model, language, що вказує на інтеграцію великих мовних моделей у взаємодію студентів із навчальними платформами.



Рис. 7. Тема 4: «Роль великих мовних моделей у навчанні»

– «Студентоцентроване навчання та персоналізація освітніх технологій» (рис. 8) зосереджується на індивідуальному підході до навчання з використанням штучного інтелекту. Найчастотніші слова *student, learn, education, study, teaching, artificial, teacher, intelligence, learning, course* свідчать про дослідження персоналізованих освітніх методик, що адаптують навчальні курси до потреб студентів. AIED у цій темі розглядається як інструмент, що може покращити навчальний досвід унаслідок адаптації програмного матеріалу, автоматизованого оцінювання та інтелектуального зворотного зв'язку.



Рис. 8. Тема 5: «Студентоцентроване навчання та персоналізація освітніх технологій»

Аналіз динаміки популярності тем

Динаміка популярності тем, виявлених за допомогою тематичного моделювання, свідчить про значні структурні зрушення в науковому дискурсі щодо інтеграції систем штучного інтелекту в освіті (AIED) у різні галузі знань (рис. 9). Зокрема теми «Штучний інтелект у медичній освіті», «Штучний інтелект в освіті» та «Студентоцентроване навчання та персоналізація освітніх технологій» демонструють експоненціальне зростання, що свідчить про збільшення інтересу до використання інтелектуальних систем для вдосконалення традиційних підходів у меди-

чній та освітній практиці. Цей тренд демонструє перехід від класичних методологій до інноваційних технологічних рішень, де AIED є каталізатором трансформації освітніх процесів і забезпечення високих етичних стандартів у підготовці медичних кадрів.

На противагу, тема «Алгоритмічні методи та машинне навчання в освіті» досягла піку наукової активності 2022 року, після чого спостерігається певне зниження публікаційної активності. Ця тенденція може вказувати на тимчасову насиченість дослідницького поля або на переорієнтацію наукового інтересу до суміжних дисциплін (наприклад, «Роль великих мовних моделей у навчанні»), що відкривають нові перспективи для розроблення інтерактивних і адаптивних навчальних систем.

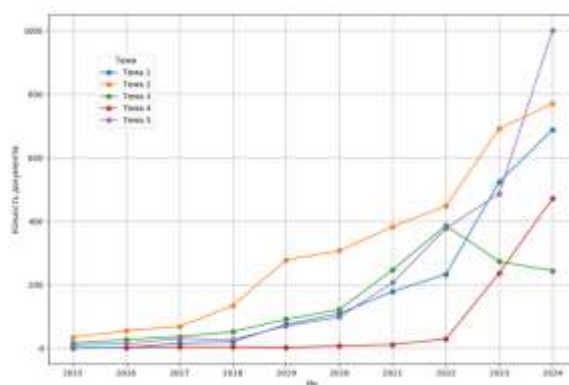


Рис. 9. Динаміка популярності тем

Отже, досягнуті результати не лише визначають актуальні наукові тренди, але й формують теоретичну основу для стратегічного планування подальших досліджень у галузі штучного інтелекту в освіті (AIED), зосереджуючись на найбільш перспективних напрямках наукового розвитку.

Висновки

Інтеграція інтелектуальних систем штучного інтелекту в освітній процес сприяє радикальній трансформації як традиційних, так і сучасних педагогічних підходів. Особливу увагу приділено студентоцентрованому навчанню та персоналізації освітніх технологій, що демонструють надзвичайно швидку експоненціальну динаміку, зокрема 2024 року цей напрям виявився найпопулярнішим серед дослідницьких тематик. Комплексний аналіз із застосуванням методів тематичного моделювання, мережевого аналізу та регресійного прогнозування свідчить про те, що адаптивні та індивідуалізовані стратегії ви-

кладання є потужним каталізатором структурних змін у системах освіти. Досягнуті результати вказують на необхідність подальшої оптимізації алгоритмічних механізмів, розроблення універсальних методологій з метою об'єктивного оцінювання ефективності впровадження ІІІ, а також посилення етичного та нормативного регулювання в цій сфері, що створює міцну базу для формування інноваційних освітніх парадигм у цифровому суспільстві.

Література

1. Бобро Н. Переваги та недоліки упровадження штучного інтелекту у освітній процес. Молодий вчений. 2024. № 4 (128). С. 72–76.
2. International Artificial Intelligence in Education Society. About the International AIED Society. URL: <https://iaied.org/about> (дата звернення 05.03.2025).
3. Nejdl W., Rensing C., Caballé S., Larriba-Pey J.L. Artificial Intelligence in Education: Challenges and Opportunities. European School Education Platform.
4. Rizzo G., De Pietro G., De Pietro L. Artificial Intelligence in Education: A Review. *Technics and Informatics in Education*. 2022. № 22. P. 223–228.
5. Dillenbourg P., Kasyanov M. M., Dillenbourg P. Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. *Le Monde*. 2024.
6. Holmes W., Bialik M., Fadel C. Artificial Intelligence: Challenges for Education and Research. *Nursing Science Quarterly*. 2023. Vol. 36, № 3. P. 227–229.
7. Blei D.M., Ng A.Y., Jordan M.I. Latent dirichlet allocation. *Journal of Machine Learning Research*. 2003. Vol. 3. P. 993–1022.
8. Schwarz C. ldagibbs: A command for topic modeling in Stata using latent Dirichlet allocation. *The Stata Journal*. 2018. Vol. 18, № 1. P. 101–117.
9. Egger R. Topic modelling. Modelling hidden semantic structures in textual data. *Applied Data Science in Tourism. Interdisciplinary Approaches, Methodologies and Applications* / за ред. R. Egger. Berlin: Springer, 2022. С. 18.
10. Obadimu A., Mead E., Agarwal N. Identifying latent toxic features on YouTube using non-negative matrix factorization. *The Ninth International Conference on Social Media Technologies, Communication, and Informatics (Valencia, 2019)*. P. 1–6.
11. Chen Y., Zhang H., Liu R., Ye Z., Lin J. Experimental explorations on short text topic mining between LDA and NMF based Schemes. *Knowledge-Based Systems*. 2019. Vol. 163. P. 1–13.
12. Lee D. D., Seung H. S. Learning the parts of objects by non-negative matrix factorization.

Nature. 1999. Vol. 401. С. 788–791.

References

1. Bobro N. (2024). Perevahy ta nedoliky uprovadzhennia shturnoho intelektu u osvittii protses [Advantages and disadvantages of implementing artificial intelligence in the educational process]. *Molodyi vchenyi*, 4 (128), 72–76. [in Ukrainian].
2. International Artificial Intelligence in Education Society. (n.d.). About the International AIED Society. Retrieved from: <https://iaied.org/about> (accessed: 05.03.2025).
3. Nejdl W., Rensing C., Caballé S., Larriba-Pey J.L. (n.d.). Artificial Intelligence in Education: Challenges and Opportunities. European School Education Platform.
4. Rizzo G., De Pietro G., De Pietro L. (2022). Artificial Intelligence in Education: A Review. *Technics and Informatics in Education*, 22, 223–228.
5. Dillenbourg P., Kasyanov M.M., Dillenbourg P. (2024). Artificial Intelligence in Education: Promises and Implications for Teaching and Learning. *Le Monde*.
6. Holmes W., Bialik M., Fadel C. (2023). Artificial Intelligence: Challenges for Education and Research. *Nursing Science Quarterly*, 36(3), 227–229.
7. Blei D.M., Ng A.Y., Jordan M.I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3(Jan), 993–1022.
8. Schwarz C. (2018). ldagibbs: A command for topic modeling in Stata using latent Dirichlet allocation. *The Stata Journal*, 18(1), 101–117.
9. Egger R. (2022). Topic modelling. Modelling hidden semantic structures in textual data. In R. Egger (Ed.), *Applied Data Science in Tourism. Interdisciplinary Approaches, Methodologies and Applications* (p. 18). Berlin: Springer.
10. Obadimu A., Mead E., Agarwal N. (2019). Identifying latent toxic features on YouTube using non-negative matrix factorization. *The Ninth International Conference on Social Media Technologies, Communication, and Informatics (Valencia)*, 1–6.
11. Chen Y., Zhang H., Liu R., Ye Z., Lin J. (2019). Experimental explorations on short text topic mining between LDA and NMF based Schemes. *Knowledge-Based Systems*, 163, 1–13.
12. Lee D.D., Seung H.S. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, 401, 788–791.

Філь Наталія Юріївна¹, к.т.н., доц. каф. автоматизації та комп'ютерно-інтегрованих технологій, тел. +38 068-617-76-94, fnu@khadi.kharkov.ua.

Слісаренко Роман Валерійович², асп. каф. медіасистем та технологій, тел. +38 066-266-09-99, roman.slisarenko@nure.ua.

Дейнеко Жанна Валентинівна², к.т.н., доц. каф. медіасистем та технологій, завідувач кафедри медіасистем та технологій, тел. +38 050-303-26-77, zhanna.deineko@nure.ua.

Морозова Лана Юріївна², ст. викл. каф. медіа-систем та технологій, лана.morozova@nure.ua.

¹Харківський національний автомобільно-дорожній університет, 61002, Україна, м. Харків, вул. Ярослава Мудрого, 25.

²Харківський національний університет радіоелектроніки, 61166, Україна, м. Харків, пр. Науки, 14.

Trends in Artificial Intelligence Research on Education: Topic Modeling Using Latent Dirichlet Allocation

Abstract. Problem. The rapid advancement of artificial intelligence (AI) has significantly impacted various domains, particularly education. AI-driven tools are increasingly being integrated into teaching methodologies, assessment frameworks, and curriculum design. However, despite the growing research interest in Artificial Intelligence in Education (AIED), the field remains fragmented, lacking a comprehensive analysis of dominant research themes and emerging trends. Understanding these trends is crucial for optimizing AI applications in education and shaping future developments in the field. **Goal.** The primary goal of this study is to analyze the current landscape of AIED research by identifying key thematic clusters, evaluating dominant research directions, and predicting future trajectories in AI-driven education. **Methodology.** This study employs network analysis, topic modeling, and global research trend evaluation to analyze scientific publications in AIED. The Latent Dirichlet Allocation (LDA) algorithm is applied to classify research articles into thematic categories, enabling the identification of primary topics shaping the field. Data is sourced from the Web of Science database, and hierarchical clustering techniques are used to determine semantic similarities between thematic models. **Results.** The analysis reveals five major research clusters in AIED, including AI applications in medical education, machine learning in education, adaptive learning technologies, large language models in learning, and student-centered AI-driven educational strategies. The study identifies increasing trends in the adoption of generative AI and chatbot-based learning support, while tradition-

al AI-driven assessment methodologies show moderate growth. The results also highlight strong interconnections between AI-driven personalization and intelligent tutoring systems, emphasizing the shift toward adaptive and student-centric learning environments. **Originality.** This research provides a structured and systematic analysis of AIED trends, leveraging advanced topic modeling techniques. Unlike previous studies focusing on isolated AI applications in education, this study presents a holistic view of thematic structures in AIED research, offering a quantitative and data-driven approach to understanding the evolution of the field. **Practical Value.** The findings can serve as a strategic reference for educators, policymakers, and AI developers in shaping future educational technologies. By identifying key thematic directions, this research supports evidence-based decision-making for integrating AI into educational systems. The results can also guide further research in AIED, helping scholars explore emerging AI-driven teaching methodologies, adaptive learning models, and ethical considerations in AI-powered education.

Key words: academic integrity, education, personalization, machine learning, latent Dirichlet allocation.

Fil Nataliia¹, candidate of technical sciences, Associate professor of the Department of Automation and Computer-Integrated Technologies, ORCID: 0000-0002-1091-1444, fnu@khadi.kharkov.ua.

Slisarenko Roman², PhD student, Department of Media Systems and Technologies, ORCID: 0009-0009-3286-4333, roman.slisarenko@nure.ua.

Deineko Zhanna², PhD in Technical Sciences, Associate Professor, Head of the Department of Media Systems and Technologies, ORCID: 0000-0003-0175-4181, zhanna.deineko@nure.ua.

Morozova Lana², Senior lecturer at the Department of Media Systems and Technologies, ORCID: 0000-0001-5317-1813, лана.morozova@nure.ua.

¹Kharkiv National Automobile and Highway University, 25, Yaroslava Mudrogo str., Kharkiv, 61002, Ukraine.

²Kharkiv National University of Radio Electronics, 14 Nauky Ave., Kharkiv, 61166, Ukraine.