

УДК 004.85:004.62

Савчук О. І., магістрант

Житомирський державний університет імені Івана Франка

СУЧАСНІ ПІДХОДИ ТА ВИКЛИКИ У ВИКОРИСТАННІ МАШИННОГО НАВЧАННЯ ДЛЯ АНАЛІЗУ ДАНИХ

В умовах сучасного цифрового суспільства більшість користувачів взаємодіє з технологіями, які узагальнено називають штучний інтелект, що не повною мірою відображає складність і різноманітність всіх можливостей цих технологій. Наприклад, популярні системи, такі як: ChatGPT, Gemini, Claude, Grok та інші, належать до класу великих мовних моделей (LLM, Large Language Models), – які є лише одним із підвидів технологій у межах машинного навчання (machine learning). Їх функціонування ґрунтується на відповідних алгоритмах і принципах, що є логічним продовженням розвитку підходів машинного навчання. Водночас LLM становлять лише частину широкого спектра сучасних рішень у галузі штучного інтелекту

Машинне навчання виступає складовою ширшої концепції штучного інтелекту та охоплює різноманітні методи і підходи. Однією з ключових сфер його застосування є аналіз великих обсягів даних, де автоматизовані алгоритми дозволяють ефективно обробляти інформацію, яка є надто масштабною для ручного опрацювання.

Практичне застосування машинного навчання в аналізі даних базується на використанні конкретних алгоритмів, кожен із яких орієнтований на вирішення певного класу задач. Залежно від мети аналізу (класифікації, прогнозування чи виявлення закономірностей) обирається відповідний підхід і модель.

Цей вибір безпосередньо залежить від формату моделі. Кожен формат має свої специфіки та розгалуження, до прикладу, різні варіанти «навчання» моделі: навчання без учителя широко використовується для виявлення прихованих закономірностей у даних, тоді як навчання з учителем – для класифікації та прогнозування. У межах задач класифікації використовуються різні алгоритмічні підходи, зокрема логістична регресія, дерева рішень і метод k-найближчих сусідів. Практичне застосування цих алгоритмів охоплює, наприклад, визначення чи є лист спамом, оцінювання кредитного ризику клієнтів. Це дозволяє не лише підвищити ефективність, але й отримати конкурентну перевагу на ринку.

Однак разом із перевагами існують і суттєві проблеми, виклики. Один із ключових – це якість даних, адже результати роботи моделей безпосередньо залежать від даних, на яких вони навчаються. Неповні,

зашумлені або упереджені дані [1] можуть призвести до некоректних результатів. Алгоритми можуть відтворювати або навіть посилювати існуючі соціальні думки, якщо вони присутні в навчальних даних [3]. Тому розробники повинні враховувати аспекти справедливості, прозорості та відповідальності при створенні систем.

Додатковою проблемою є обмежена інтерпретованість складних моделей, зокрема глибоких нейронних мереж (DNN, Deep Neural Network) і великих мовних моделей. Їх функціонування часто є непрозорим для користувача, що ускладнює розуміння механізмів отримання результатів, а також оцінювання їхньої достовірності [2]. Це створює труднощі у довірі до таких систем та їх впровадженні в критично важливі процеси.

Також варто згадати про технічні обмеження, зокрема високі вимоги до обчислювальних ресурсів. Навчання складних моделей потребує значних потужностей, що може бути недоступним для малих компаній або окремих дослідників. Разом з цим, збільшення потужностей прямопропорційно впливає на збільшення витрат підтримки життєдіяльності систем штучного інтелекту, що з часом може критично вплинути на навколишнє середовище [4].

Попри це, перспективи машинного навчання залишаються значними, а його інтеграція в різні сфери відкриває нові можливості для розвитку. Водночас ключовим є забезпечення відповідального та обґрунтованого використання цих технологій. Отже, сучасні підходи до застосування машинного навчання демонструють суттєвий прогрес, але потребують комплексного врахування як технічних, так і етичних аспектів для ефективного використання їх потенціалу.

Список використаних джерел:

1. Data Cleaning. Definition, types, and examples. Julius. URL: <https://julius.ai/glossary/data-cleaning> (дата звернення: 22.03.2026).
2. Columbia Journalism Review. AI Search Has a Citation Problem. Columbia Journalism Review. URL: https://www.cjr.org/tow_center/we-compared-eight-ai-search-engines-theyre-all-bad-at-citing-news.php (дата звернення: 22.03.2026).
3. What is AI bias? | SAP – System Analysis Program Development. URL: <https://www.sap.com/resources/what-is-ai-bias#top> (дата звернення: 22.03.2026).
4. Science Feedback. How much electricity do data centers use, and can they run on renewables?. Science Feedback. URL: <https://cutt.ly/ytl6B4W3> (дата звернення: 22.03.2026).