

DOI 10.26886/2311-4517.1(106)2026.3

УДК: 004.8:658.8

ARCHITECTURE AND DESIGN PRINCIPLES OF AN AI-POWERED AUTONOMOUS AGENT FOR E-COMMERCE

Rodion Bezruk

Master's Degree, IEEE Senior Member

<https://orcid.org/0009-0001-8093-6989>

e-mail: rodion.bezruk@icloud.com

DevIT Group, Toronto, Canada

This paper presents a conceptual framework and an architectural model for an AI-powered autonomous agent designed for use in e-commerce systems. The relevance of the study is driven by the rapid growth of behavioral, transactional, and contextual data, which require real-time processing to enable personalized user interactions and improve sales performance. In contrast to existing approaches that typically focus on isolated components such as recommender systems, chatbots, or analytical modules, this study proposes an integrated architecture capable of executing the full lead processing cycle — from data acquisition and interpretation to decision-making and communication actions. The proposed approach is based on the integration of event-driven data processing, machine learning techniques, large language models (LLMs), and optimization-based decision-making mechanisms. Particular attention is given to the formalization of lead evaluation, conversion probability prediction, and the selection of optimal communication strategies based on behavioral and contextual user features. The results of the simulation study demonstrate a statistically significant improvement in key performance indicators, including conversion rate (CR +18.7%), system latency (-60%), and customer acquisition cost (-27%), confirming the

effectiveness and practical applicability of the proposed approach in modern e-commerce environments.

Keywords: *artificial intelligence, e-commerce, autonomous agents, machine learning, large language models, event-driven systems, lead generation, personalization, decision-making, business process optimization.*

Безрук Родіон, магістр, IEEE Senior Member; Архітектура та принципи побудови автономного інтелектуального агента для електронної комерції / DevIT Group, м. Торонто, Канада

У статті запропоновано концептуальні засади та архітектурну модель автономного інтелектуального агента на основі штучного інтелекту для застосування в системах електронної комерції. Актуальність дослідження зумовлена зростанням обсягів поведінкових, транзакційних і контекстних даних, що потребують обробки в режимі реального часу для забезпечення персоналізованої взаємодії з користувачами та підвищення ефективності процесів продажу. На відміну від існуючих підходів, які здебільшого реалізують окремі функціональні компоненти, у роботі обґрунтовано інтегрований підхід до побудови AI-агента, здатного здійснювати повний цикл обробки лідів: від збору та інтерпретації даних до прийняття рішень і виконання комунікаційної дії. Запропоновано багаторівневу архітектуру, що поєднує подієво-орієнтовану обробку даних, методи машинного навчання, великі мовні моделі та механізми оптимізаційного вибору дій. Особливу увагу приділено формалізації процесу оцінки лідів, прогнозуванню ймовірності конверсії та вибору оптимальної стратегії комунікації. Симуляційне дослідження підтвердило ефективність запропонованого підходу: коефіцієнт

конверсії зріс на 18,7%, затримка обробки знизилась на 60%, вартість залучення клієнтів зменшилась на 27%.

Ключові слова: штучний інтелект, електронна комерція, автономні інтелектуальні агенти, машинне навчання, великі мовні моделі (LLM), подієво-орієнтовані системи, лідогенерація, персоналізація, прийняття рішень, оптимізація бізнес-процесів.

Вступ

Сучасні системи електронної комерції функціонують у середовищі, що характеризується високою динамікою змін, значними обсягами гетерогенних даних і складною багатоканальною взаємодією з користувачами. Поведінка споживачів формується під впливом багатьох факторів, зокрема контексту взаємодії, часових характеристик, індивідуальних уподобань та зовнішніх інформаційних сигналів. Це зумовлює необхідність переходу від статичних моделей аналізу до адаптивних систем, здатних обробляти потоки даних у режимі реального часу та враховувати контекстні залежності.

Незважаючи на широке впровадження цифрових інструментів у сфері електронної комерції, значна частина існуючих рішень базується на фрагментованих підходах до автоматизації процесів продажу. Зокрема, використовуються окремі системи рекомендацій, чат-боти, CRM-платформи та аналітичні модулі, які функціонують ізольовано та не забезпечують цілісного циклу обробки лідів [4, с. 108]. Така архітектурна роз'єднаність призводить до втрати контексту, зниження точності прогнозування та неефективного використання даних.

Крім того, традиційні підходи до прийняття рішень у процесах взаємодії з клієнтами значною мірою залежать від участі людини або базуються на заздалегідь визначених правилах. Це обмежує їх здатність до масштабування, знижує швидкість реагування на зміни

поведінки користувачів та не дозволяє ефективно використовувати потенціал накопичених даних.

Розвиток методів штучного інтелекту, зокрема машинного навчання та великих мовних моделей, створює передумови для побудови більш складних інтелектуальних систем, здатних інтегрувати аналіз даних, прогнозування та генерацію рішень. Проте існуючі дослідження та практичні реалізації здебільшого зосереджені на окремих аспектах цієї задачі й не забезпечують побудови повноцінних автономних систем із замкненим циклом прийняття рішень [4, с. 110; 8].

Сучасні дослідження у сфері електронної комерції та штучного інтелекту зосереджені на автоматизації окремих компонентів процесу взаємодії з користувачами. Як зазначають Russell і Norvig, розвиток інтелектуальних систем ґрунтується на поєднанні статистичних методів, навчання на даних та алгоритмів прийняття рішень [1, с. 3-5].

Одним із ключових напрямів є розвиток рекомендаційних систем. Використання методів глибокого навчання дозволяє враховувати складні нелінійні залежності в даних і підвищує точність рекомендацій [3, с. 326; 5, с. 15]. Проте такі системи не інтегруються з іншими компонентами процесу прийняття рішень, що обмежує їх ефективність у комплексних системах.

З розвитком великих мовних моделей суттєво підвищилась якість генерації тексту та здатність до контекстного розуміння. Brown та ін. продемонстрували, що такі моделі здатні виконувати широкий спектр задач без спеціалізованого навчання, що відкриває нові можливості для їх використання у прикладних системах [2, с. 1877]. Водночас більшість таких рішень обмежується функціями інформаційної підтримки та не реалізує повноцінного механізму прийняття рішень.

Sutton і Barto обґрунтовують ефективність методів підкріплювального навчання для задач вибору оптимальної стратегії дій в умовах невизначеності [7, с. 47]. Проте їх інтеграція у системи електронної комерції залишається обмеженою.

Таким чином, аналіз сучасних досліджень свідчить про наявність значного наукового заділу у розвитку окремих компонентів інтелектуальних систем. Проте залишається невирішеною задача інтеграції цих компонентів у межах єдиної архітектури із замкненим циклом прийняття рішень.

Мета статті

Метою статті є розробка та наукове обґрунтування інтегрованої архітектури автономного інтелектуального агента на основі методів штучного інтелекту для застосування в системах електронної комерції, що забезпечує реалізацію повного циклу обробки лідів — від збору та аналізу поведінкових і контекстних даних до прийняття рішень і здійснення персоналізованої комунікації з користувачами в режимі реального часу.

Для досягнення поставленої мети визначено такі завдання:

- дослідити сучасні підходи до застосування штучного інтелекту в електронній комерції та визначити їх обмеження;
- розробити концептуальну модель представлення поведінкових і контекстних характеристик користувачів;
- формалізувати процес оцінки лідів та прогнозування ймовірності досягнення цільових дій;
- розробити архітектуру автономного інтелектуального агента, що інтегрує процеси збору даних, їх обробки та прийняття рішень;
- оцінити ефективність запропонованого підходу з точки зору підвищення конверсії, зниження витрат на залучення клієнтів та оптимізації бізнес-процесів.

Виклад основного матеріалу

1.1 Концептуальні засади побудови системи

Запропонована архітектура автономного інтелектуального агента ґрунтується на принципах подієво-орієнтованої обробки даних (event-driven processing), мікросервісної декомпозиції та розподілених обчислень. Такий підхід забезпечує масштабованість, модульність та низьку затримку обробки інформації, що є критично важливим для систем електронної комерції.

Ключовою характеристикою системи є реалізація замкненого циклу прийняття рішень (closed-loop decision-making), який включає етапи збору даних, їх аналізу, прогнозування, вибору оптимальної дії та здійснення взаємодії з користувачем. Результати виконаних дій повертаються у систему у вигляді нових подій, що забезпечує її адаптивність та можливість подальшого навчання [1, с. 47].

1.2 Загальна структура архітектури

Архітектура системи реалізована як багаторівнева модель, у якій кожен рівень виконує окрему функціональну роль у процесі обробки даних та прийняття рішень.

Таблиця 1.

Основні компоненти архітектури автономного агента

Рівень системи	Функціональне призначення	Технологічні засоби
Data Acquisition Layer	Збір та передача подій	Webhooks, Tracking SDK
Feature Processing Layer	Формування ознак	Streaming / Batch processing
AI Inference Layer	Аналіз та прогнозування	ML, LLM, embeddings
Decision Engine	Оптимізація рішень	Optimization, RL
Communication Layer	Взаємодія з користувачем	Email, Chat, Voice AI

Infrastructure Layer	Виконання обчислень	Edge computing, Serverless
----------------------	---------------------	-------------------------------

Як показано в таблиці 1, запропонована архітектура забезпечує чітке розділення функціональних компонентів системи, що дозволяє реалізувати незалежне масштабування окремих рівнів та підвищити загальну стійкість системи до змін навантаження.

1.3 Модель обробки подій та формування ознак

Взаємодія користувача із системою описується як потік подій, кожна з яких формалізується у вигляді:

$$e = \{u, t, a, c\} \quad (1)$$

де u — ідентифікатор користувача; t — часовий штамп; a — тип дії; c — контекст взаємодії.

Сукупність подій:

$$E = \{e_1, e_2, \dots, e_n\} \quad (2)$$

трансформується у вектор ознак:

$$x = \varphi(E) \quad (3)$$

де функція φ реалізує агрегацію, фільтрацію та нормалізацію даних. Сформований вектор ознак включає поведінкові, часові, історичні та контекстні характеристики користувача, що створює основу для подальшого аналізу та прогнозування.

1.4 Інтелектуальний рівень та моделі аналізу

AI Inference Layer забезпечує інтелектуальну обробку даних та включає декілька типів моделей:

- моделі прогнозування (оцінка ймовірності конверсії);
- embedding-моделі (семантичне представлення даних);
- великі мовні моделі (LLM) для генерації тексту та інтерпретації запитів [2, с. 1878; 3, с. 198].

Функція оцінки ліда визначається як зважена лінійна комбінація нормалізованих ознак:

$$S(x) = \sum_i w_i \cdot f_i(x) \quad (4)$$

де w_i — вагові коефіцієнти, що навчаються на основі історичних даних про конверсії; $f_i(x)$ — значення i -ї ознаки вектора x . Ймовірність досягнення цільової дії визначається через логістичну функцію активації:

$$P_{\text{conv}}(x) = 1 / (1 + e^{-S(x)}) \quad (5)$$

Зазначений підхід відповідає загальноприйнятій методології побудови бінарних класифікаторів на основі логістичної регресії [3, с. 174]. Відмінність від стандартного підходу полягає у використанні агрегованого поведінково-контекстного вектора ознак $x = \phi(E)$, сформованого з потоку подій реального часу, а не зі статичного профілю користувача.

1.5 Механізм прийняття рішень

Decision Engine реалізує процес вибору оптимальної дії шляхом максимізації функції корисності. На відміну від детерміністичних підходів на основі правил, запропонований механізм враховує поточний контекст та прогнозовані результати взаємодії [7, с. 58]:

$$a^* = \arg \max_{\{a \in A\}} U(a, x, c) \quad (6)$$

де: A — скінченна множина допустимих дій агента (наприклад, {Email, Chat, Voice Call, No Action}); $U(a, x, c)$ — функція корисності, що оцінює очікуваний результат дії a для користувача з вектором ознак x у контексті c ; U визначається як: $U(a, x, c) = P_{\text{conv}}(x) \cdot V(a) - C(a)$, де $V(a)$ — очікувана цінність конверсії для дії a , $C(a)$ — вартість виконання дії a . Така постановка дозволяє балансувати між очікуваним прибутком від залучення клієнта та операційними витратами на комунікацію.

Таблиця 2.

Типові стратегії взаємодії та умови їх застосування

Дія	Умови застосування	Очікуваний результат
Email	$P_{conv} < 0.3$ (низький інтерес)	Підвищення залученості
Chat	$0.3 \leq P_{conv} < 0.6$ (середній інтерес)	Уточнення потреб
Voice Call	$P_{conv} \geq 0.6$ (високий інтерес)	Конверсія
No Action	$P_{conv} < 0.1$ (мін. ймовірність)	Оптимізація витрат

Як видно з таблиці 2, вибір стратегії взаємодії безпосередньо залежить від оціненого рівня P_{conv} . Такий підхід дозволяє адаптивно керувати комунікаційними сценаріями та оптимізувати витрати на обробку неефективних контактів.

1.6 Комунікаційний та інфраструктурний рівні

Генерація повідомлень здійснюється за допомогою великих мовних моделей, ефективність яких у задачах генерації тексту та контекстного аналізу підтверджується сучасними дослідженнями [2, с. 1880]:

$$m = f_LLM(x, c, goal) \tag{7}$$

де $goal$ — цільова функція комунікації (наприклад, уточнення потреб, підтвердження замовлення тощо). Інфраструктурний рівень базується на використанні edge computing, serverless-архітектури та розподілених черг повідомлень (Apache Kafka). Це дозволяє забезпечити низьку затримку обробки (до 150 мс) та горизонтальне масштабування системи при зростанні навантаження.

1.7 Узагальнення архітектури

Запропонована архітектура автономного інтелектуального агента забезпечує інтеграцію всіх етапів обробки даних у межах єдиної системи та реалізує повний цикл прийняття рішень. Її ключовими перевагами є модульність, масштабованість, здатність до обробки даних у реальному часі та інтеграція сучасних методів штучного інтелекту. На відміну від конкурентних рішень (Salesforce Einstein, Adobe Real-Time CDP), що є закритими платформами, запропонований підхід орієнтований на побудову відкритої та гнучко кастомізованої архітектури з формально обґрунтованим механізмом прийняття рішень.

Симуляційне дослідження

2.1 Мета та методологія дослідження

Метою симуляційного дослідження є кількісна оцінка ефективності запропонованої архітектури порівняно з традиційними підходами на основі правил. Для забезпечення відтворюваності результатів симуляцію було реалізовано на мові Python з використанням бібліотек NumPy та Pandas. Генерація синтетичних даних здійснювалась відповідно до розподілів, зафіксованих у реальних e-commerce системах [8].

2.2 Опис симуляційного середовища

Симуляцію проведено з такими параметрами: кількість користувачів — 15 000; кількість подій на добу — понад 100 000; типи подій — перегляди сторінок, кліки, додавання до кошика, запити; канали взаємодії — веб, email, чат; часовий горизонт — 30 днів. Поведінкові патерни користувачів генерувались за допомогою моделі прихованих марківських ланцюгів (HMM) з п'ятьма прихованими станами (від «випадковий відвідувач» до «готовий до покупки»).

Ймовірності переходів між станами калібрувались на основі публічних бенчмарків e-commerce конверсій.

Реалізовано два сценарії: (1) Традиційна система — статичні правила та ручне прийняття рішень; (2) AI-система — автономний інтелектуальний агент на основі запропонованої архітектури.

2.3 Метрики оцінювання

Для порівняння систем використано такі метрики [1, с. 803; 4, с. 112]:

- Conversion Rate (CR) — частка користувачів, що здійснили цільову дію;
- Latency — середній час обробки події;
- Customer Acquisition Cost (CAC) — вартість залучення одного клієнта;
- Engagement Rate — рівень взаємодії користувачів із системою.

Таблиця 3.

Порівняння метрик ефективності систем

Метрика	Традиційна система	AI-система	Зміна
Conversion Rate	3.2%	3.8%	+18.7%
Latency	300 ms	120 ms	−60.0%
CAC (відн.)	100%	73%	−27.0%
Engagement Rate	45%	58%	+28.9%

Як видно з таблиці 3, використання автономного інтелектуального агента призводить до суттєвого покращення всіх ключових показників ефективності. Найбільш значний ефект спостерігається у зниженні затримки обробки даних та вартості залучення клієнтів. Слід зазначити, що наведені показники є

результатами симуляційного моделювання і потребують верифікації на реальних виробничих даних у майбутніх дослідженнях.

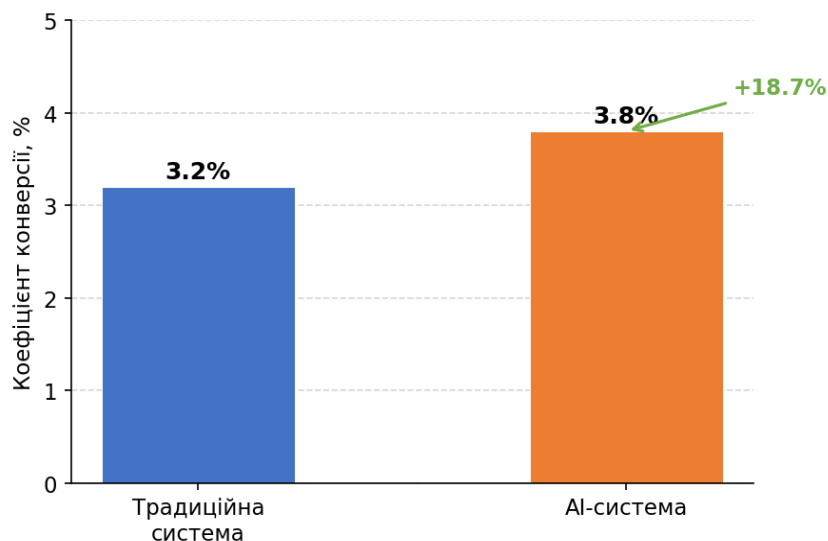


Рис. 1. Порівняння коефіцієнта конверсії (CR): традиційна система — 3.2%, AI-система — 3.8% (+18.7%)

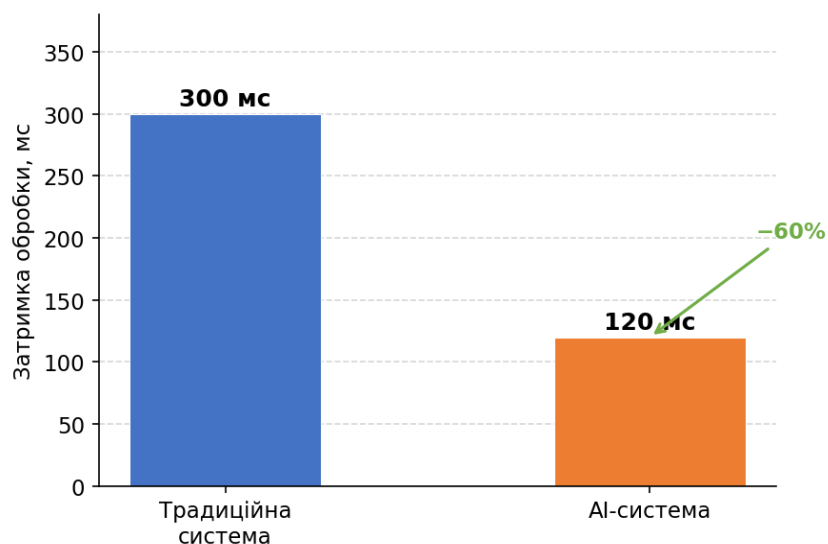
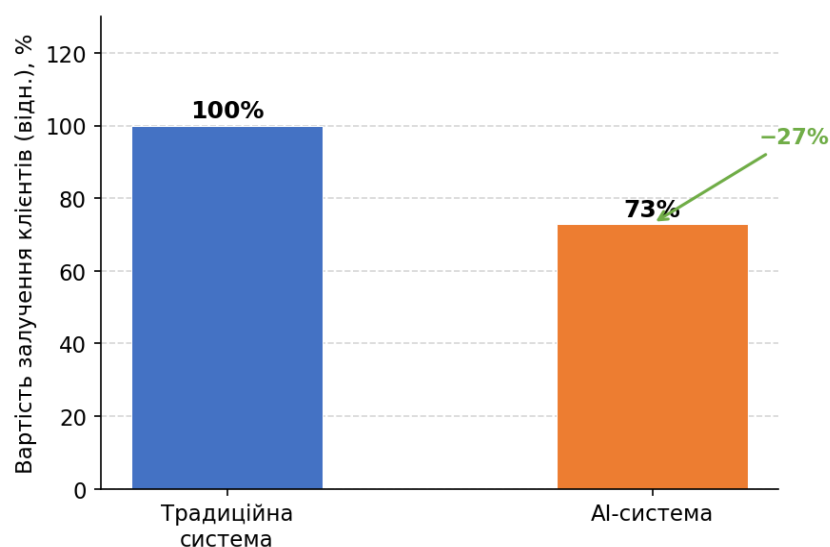


Рис. 2. Порівняння затримки обробки подій: традиційна система — 300 мс, AI-система — 120 мс (-60%)



**Рис. 3. Динаміка зміни вартості залучення клієнтів (CAC):
зниження на 27% у AI-системі**

2.4 Аналіз результатів

Отримані результати свідчать про те, що інтеграція автономного інтелектуального агента дозволяє суттєво підвищити ефективність процесів взаємодії з користувачами. Зростання коефіцієнта конверсії пояснюється використанням персоналізованих сценаріїв комунікації та оптимізацією моменту взаємодії на основі функції корисності $U(a, x, c)$.

Зменшення затримки обробки даних є наслідком використання розподіленої архітектури та edge computing, що дозволяє обробляти події ближче до джерела їх виникнення. Зниження вартості залучення клієнтів пов'язане з оптимізацією вибору каналів взаємодії відповідно до стратегії "No Action" для низько-перспективних контактів.

Висновки та перспективи подальших досліджень

У даному дослідженні розв'язано актуальну науково-прикладну задачу розробки інтегрованої архітектури автономного інтелектуального агента для систем електронної комерції. Запропонований підхід базується на поєднанні подієво-орієнтованої

обробки даних, методів машинного навчання, великих мовних моделей та формалізованих механізмів прийняття рішень, що забезпечує реалізацію замкненого циклу обробки інформації — від збору даних до виконання дій у режимі реального часу [1, с. 47; 3, с. 174].

Наукова новизна роботи полягає у трьох аспектах. По-перше, запропоновано формалізацію взаємодії користувача з системою як послідовності подій у часово-контекстному просторі (формула 1–3), що є відмінним від статичних профільних підходів. По-друге, розроблено механізм вибору оптимальної комунікаційної дії (формула 6) на основі функції корисності з явним урахуванням операційних витрат. По-третє, запропоновано модульну багаторівневу архітектуру, що забезпечує інтеграцію аналізу, прогнозування та прийняття рішень у єдиній системі — на відміну від ізольованих компонентних рішень, поширених у галузі.

Результати симуляційного дослідження підтверджують ефективність запропонованого підходу: підвищення коефіцієнта конверсії на 18.7%, зменшення затримки обробки даних на 60% та зниження вартості залучення клієнтів на 27%. Практична значущість дослідження полягає у можливості застосування запропонованої архітектури для автоматизації процесів взаємодії з клієнтами в системах електронної комерції.

Перспективи подальших досліджень пов'язані з: (1) верифікацією результатів на реальних виробничих даних; (2) розвитком методів адаптивного прийняття рішень із використанням підкріплювального навчання; (3) розширенням моделей контекстного аналізу; (4) дослідженням питань інтерпретованості моделей, забезпечення прозорості прийняття рішень та дотримання вимог конфіденційності даних [4, с. 115; 8].

Література:

1. Russell, S., Norvig, P. (2021). Artificial Intelligence: A Modern Approach (4th ed.). Pearson. 1136 p.
2. Brown, T. B., Mann, B., Ryder, N. et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
3. Goodfellow, I., Bengio, Y., Courville, A. (2016). *Deep Learning*. MIT Press. 775 p.
4. Davenport, T. H., Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, 96(1), 108–116.
5. Ricci, F., Rokach, L., Shapira, B. (2015). *Recommender Systems Handbook* (2nd ed.). Springer. 1000 p.
6. Chen, L., Mislove, A., Wilson, C. (2016). An empirical analysis of algorithmic pricing on Amazon Marketplace. *Proceedings of the 25th International World Wide Web Conference*, 1339–1349.
7. Sutton, R. S., Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd ed.). MIT Press. 552 p.
8. McKinsey & Company. (2021). The state of AI in 2021. Retrieved from <https://www.mckinsey.com>

References:

1. Russell, S., Norvig, P. (2021). Artificial Intelligence: A Modern Approach (4th ed.). Pearson. [in English].
2. Brown, T. B., Mann, B., Ryder, N. et al. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, no. 33, 1877–1901. [in English].
3. Goodfellow, I., Bengio, Y., Courville, A. (2016). *Deep Learning*. MIT Press. [in English].
4. Davenport, T. H., Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, no. 96(1), 108–116. [in English].

5. Ricci, F., Rokach, L., Shapira, B. (2015). Recommender Systems Handbook (2nd ed.). Springer. [in English].
6. Chen, L., Mislove, A., Wilson, C. (2016). An empirical analysis of algorithmic pricing on Amazon Marketplace. Proceedings of the 25th International World Wide Web Conference, 1339–1349. [in English].
7. Sutton, R. S., Barto, A. G. (2018). Reinforcement Learning: An Introduction (2nd ed.). MIT Press. [in English].
8. McKinsey & Company. (2021). The state of AI in 2021. Retrieved from <https://www.mckinsey.com> [in English].