

<https://doi.org/10.30857/2786-5371.2025.5.2>Received: 03.09.2025
Revised: 08.10.2025
Accepted: 22.10.2025

УДК 004.056.5

Ганна ЗАВГОРОДНЯ¹, Валерій ЗАВГОРОДНІЙ¹,
Ольга ТКАЧЕНКО², Андрій САВЧЕНКО³¹ Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Україна² Київський національний університет імені Тараса Шевченка, Україна³ Заклад вищої освіти «Міжнародний науково-технічний університет імені академіка Юрія Бугая», Київ, Україна**МЕТОДОЛОГІЯ РОЗРОБЛЕННЯ СИСТЕМИ
АДАПТИВНОЇ ГЕНЕРАЦІЇ КОНТЕНТУ**

Мета. Метою статті є розробка методології створення системи адаптивної генерації контенту, яка забезпечує високий рівень персоналізації, масштабованості та інтеграцію сучасних алгоритмів машинного навчання. Особлива увага приділяється поетапній обробці даних через ETL-процеси, потокову обробку у режимі реального часу з мікробатчами 1 секунда та інтеграцію ML-моделей для динамічного налаштування контенту відповідно до поведінки користувачів.

Методика. У роботі запропоновано комбінований підхід, який поєднує ETL-процеси для підготовки та нормалізації даних, потокову обробку для забезпечення реального часу та мінімізації Latency, а також інтеграцію ML-моделей з інкрементальним оновленням для прогнозування та адаптації контенту. Методика включає формалізацію системи через функціональні блоки, математичні моделі обробки даних і сучасні фреймворки для масштабування обчислень, що забезпечують гнучке управління великими потоками подій та користувацькими сценаріями.

Результати. Експериментальна перевірка проводилась на датасеті, що містить 1 200 000 подій взаємодії, 48 000 унікальних користувачів та 12 500 одиниць контенту протягом 30 днів. Порівняння потокового та пакетного підходів продемонструвало зростання Assigasy на 6 процентних пунктів, підвищення Precision, Recall та F1-score, а також зменшення Latency більш ніж удвічі. Візуальний аналіз та статистична перевірка підтвердили стабільність і передбачуваність роботи потокової архітектури, що свідчить про комплексне покращення якості персоналізації та продуктивності системи.

Наукова новизна. Запропоновано формалізовану модель системи адаптивної генерації контенту, яка об'єднує ETL-процеси, потокову обробку з мікробатчами та інтеграцію ML-моделей у єдиному архітектурному рішенні. Розроблено алгоритми інкрементального оновлення моделей, динамічного балансування навантаження та адаптивного налаштування контенту в реальному часі, що раніше не описувалося в сучасній літературі.

Практична значимість. Запропонована методологія може бути використана при створенні систем рекомендацій, онлайн-освітніх платформ, ігор та інших цифрових сервісів, де важлива персоналізація та адаптивність контенту. Вона дозволяє скоротити витрати на обробку даних, підвищити користувацький досвід, забезпечити стабільність і передбачуваність роботи системи, а також полегшити інтеграцію сучасних алгоритмів машинного навчання у корпоративні рішення.

Ключові слова: адаптивна генерація контенту; ETL-процеси; потокова обробка; машинне навчання; персоналізація; масштабовані системи.

Вступ. Сучасні цифрові платформи, включно з онлайн-освітніми системами, іграми та медіасервісами, дедалі більше покладаються на адаптивну генерацію контенту для підвищення якості взаємодії з користувачем [1, 2]. Саме для цього активно застосовуються персоналізовані системи контент-менеджменту, які автоматично формують або адаптують інформацію на основі поведінкових патернів користувачів, що підвищує ефективність цифрових сервісів нового покоління [3, 4].

Традиційні підходи до адаптації контенту, зокрема статичні алгоритми чи прості фільтраційні моделі, вже не забезпечують достатньої ефективності при обробці великих

обсягів даних і в умовах реального часу [5, 6]. У відповідь на ці виклики активно використовуються ETL-процеси (Extract, Transform, Load), які дозволяють здійснювати попередню обробку даних, їхнє очищення, інтеграцію з різних джерел і підготовку для подальшого аналітичного та програмного використання [7, 8].

Потокова обробка даних (stream processing) стала ключовою технологією у системах, що потребують обробки подій у реальному часі без затримок, властивих пакетній обробці [9, 10]. Завдяки поточковим технологіям забезпечується масштабованість і низькі затримки при високих навантаженнях, що є критичною вимогою для сучасних адаптивних систем [11].

Інтеграція моделей машинного навчання (ML) у процеси генерації та адаптації контенту дозволяє не просто обробляти дані, а прогнозувати потреби користувачів і автоматично підлаштовувати системну поведінку під індивідуальні профілі [12]. Так, дослідження західних авторів демонструють ефективність моделей у прогнозуванні поведінки та персоналізації [1, 2]. Інші роботи [13, 15, 16] показують важливість застосування нейромережових агентів для моделювання поведінки користувача та створення масштабованих архітектур для адаптивних систем.

Незважаючи на численні досягнення, досі не існує уніфікованої методології, яка б об'єднувала ETL-процеси, потокову обробку даних та інтеграцію ML-моделей у єдину адаптивну систему, здатну одночасно забезпечувати персоналізацію, масштабованість і роботу в режимі реального часу [5, 10, 14]. Такий брак системного підходу обмежує можливості практичної реалізації адаптивних генераторів контенту у реальних виробничих середовищах, де важлива як продуктивність, так і якість персоналізації [9, 12].

Отже, існує потреба в розробці структурованої методології, що поєднує ключові компоненти адаптивних систем: обробку даних, потокову обробку та машинне навчання. Саме така методологія є предметом дослідження даної статті [15, 16].

Постановка завдання. Метою даного дослідження є розробка методології створення адаптивної системи генерації контенту, яка поєднує ETL-процеси, потокову обробку даних та інтеграцію машинного навчання для забезпечення персоналізації та ефективності у режимі реального часу. Попередні підходи до автоматизації адаптивного контенту обмежувались окремими аспектами: або обробкою великих обсягів даних через ETL-фреймворки [7, 12], або використанням поточкових технологій для швидкого реагування на події [5, 9], або впровадженням моделей ML для прогнозування поведінки користувача [14, 16]. Система, що пропонується, має вирішити такі завдання:

- збір та інтеграція даних із різних джерел у єдину платформу з підтримкою потокової обробки для мінімізації затримок;
- попередня обробка даних (очищення, трансформація, нормалізація) із використанням ETL-процесів для підготовки якісної інформації до аналізу;
- інтеграція ML-моделей для адаптивного прогнозування та персоналізації контенту відповідно до поведінкових патернів користувачів;
- масштабованість та гнучкість системи для підтримки великих потоків даних та численних користувачів без втрати продуктивності;
- підвищення точності та ефективності генерації контенту через поєднання алгоритмічних та нейромережових підходів.

Розробка методології передбачає формалізацію процесів ETL, потокової обробки та ML-інтеграції як єдиного циклу роботи системи, що дозволяє забезпечити динамічну адаптацію до зміни поведінки користувача, швидке реагування на нові дані та підтримку високої якості персоналізованого контенту.

Таким чином, завдання полягає у створенні структурованої, масштабованої та автономної системи адаптивної генерації контенту, яка буде здатна працювати в умовах

потокової обробки великих обсягів даних, забезпечувати високу якість персоналізації та інтегрувати сучасні ML-моделі для прогнозування поведінки користувача.

Результати дослідження. Система адаптивної генерації контенту формально описується як кортеж:

$$S = \langle D, P, M, C, R \rangle,$$

де D – множина джерел даних, включаючи структуровані бази даних та неструктуровані веб-дані;

P – ETL-процеси (Extract, Transform, Load), що забезпечують інтеграцію, очищення та нормалізацію даних;

M – модулі машинного навчання для аналізу та прогнозування поведінки користувачів;

C – адаптивний модуль генерації контенту, що враховує результати ML-аналізу для персоналізації;

R – модуль оцінки якості та зворотного зв'язку, що дозволяє коригувати алгоритми генерації контенту.

Така інтегрована архітектура дозволяє підтримувати динамічну взаємодію всіх компонентів та забезпечує масштабованість для великих потоків даних.

ETL-процеси. ETL-процеси забезпечують збір, трансформацію та підготовку даних для подальшого аналізу. Кожен етап визначається наступним чином:

1. Extract – збір даних із різних джерел (бази даних, API, веб-ресурси).

2. Transform – очищення, нормалізація та агрегування даних. Для поточкових подій трансформація виконується у режимі near-real-time.

3. Load – завантаження підготовлених даних у сховище для використання ML-моделями.

Формально трансформацію можна записати як:

$$D_t = f(T(D_e)),$$

де D_e – необроблені дані після Extract;

$T(D_e)$ – функція трансформації (очищення, нормалізація, агрегування);

$f(T(D_e))$ – функція завантаження в сховище;

D_t – підготовлені дані для аналізу.

Потокова обробка. Потокова обробка дозволяє системі швидко реагувати на нові події та підтримувати актуальність контенту. Використовуються черги повідомлень (Kafka) та потокові платформи (Spark Streaming).

Метрики поведінки користувачів обчислюються в реальному часі:

$$F_u(t) = \frac{\sum_{i=1}^n f_i}{t}, T_u(t) = \frac{\sum_{i=1}^n \Delta t_i}{n},$$

де $F_u(t)$ – частота переглядів користувача u ;

$T_u(t)$ – середня тривалість взаємодії;

f_i – подія перегляду;

Δt_i – тривалість взаємодії;

n – кількість подій у проміжку часу t .

Інтеграція ML-моделей. Модулі машинного навчання виконують прогнозування уподобань користувачів та рекомендацію контенту. Формально поведінку користувача u можна представити як:

$$y_u = M_u(D_t, F_u, T_u, H_u),$$

де D_t – дані після ETL;
 F_u, T_u – поведінкові метрики;
 H_u – історія взаємодії користувача;
 Y_u – прогнозоване уподобання або реакція на контент.

Для підвищення точності використовується ансамблювання моделей та нейронні мережі: LSTM для послідовностей подій і CNN для аналізу мультимедійного контенту.

Модуль генерації контенту. Генерація контенту здійснюється адаптивними алгоритмами, які враховують прогноз ML-моделей:

$$C_u = g(y_u, P_c),$$

де P_c – параметри контенту;
 C_u – згенерований контент для користувача u .

Зворотний зв'язок використовується для оновлення моделей у режимі online learning:

$$M_u^{new} = h(M_u^{old}, R),$$

де R – оцінка ефективності контенту.

Експериментальні результати. Для комплексного аналізу ефективності запропонованої системи адаптивної генерації контенту було використано сукупність кількісних метрик, які охоплюють точність моделей персоналізації, продуктивність потокової обробки, затримку сервісу, надійність інфраструктури та стійкість до навантаження. Такий багатовимірний підхід дозволяє оцінити систему не лише з позиції якості рекомендацій, але й з точки зору її експлуатаційної придатності у реальних умовах.

1. Ассурасу (точність класифікації / рекомендації) визначається як частка правильно сформованих рекомендацій або коректно класифікованих подій відносно загальної кількості передбачень:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN},$$

де TP (True Positive) – релевантні рекомендації, з якими користувач взаємодівав;
 TN (True Negative) – коректно відхилені нерелевантні об'єкти;
 FP (False Positive) – нерелевантний контент, який був рекомендований;
 FN (False Negative) – релевантний контент, який система не запропонувала.

У контексті адаптивної генерації контенту ця метрика відображає загальний рівень адекватності персоналізації. Водночас слід зазначити, що для систем рекомендацій Ассурасу не завжди є вичерпною, особливо при дисбалансі класів. Саме тому додатково можуть використовуватися Precision, Recall та F1-score. Проте у межах даного дослідження Ассурасу застосовано як інтегральний показник узагальненої якості моделі.

Зростання Ассурасу у потоковій конфігурації пояснюється оперативним оновленням моделей на основі актуальних поведінкових даних, що мінімізує ефект «застарілих» профілів користувачів.

2. Precision (точність позитивних прогнозів) характеризує частку релевантного контенту серед усіх запропонованих рекомендацій:

$$Precision = \frac{TP}{TP+FP}.$$

Ця метрика особливо важлива для персоналізованих сервісів, оскільки відображає ступінь «чистоти» рекомендаційного списку. Низьке значення Precision означає переваження користувача нерелевантним контентом, що може призвести до зниження довіри до системи.

У рамках експерименту було зафіксовано підвищення Precision після впровадження потокового оновлення моделей, що підтверджує зменшення кількості хибнопозитивних рекомендацій.

3. Recall (повнота виявлення релевантного контенту) визначає здатність системи знаходити всі потенційно релевантні об'єкти:

$$Recall = \frac{TP}{TP+FN}$$

Для систем адаптивної генерації контенту високе значення Recall означає, що система не «втрачає» значущі інтереси користувача. Однак між Precision та Recall існує баланс: підвищення одного показника часто зменшує інший. Тому у практичному застосуванні оптимізується компромісне співвідношення між ними.

4. F1-score є гармонійним середнім між Precision та Recall:

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Ця метрика використовується як збалансований індикатор якості моделі, особливо в умовах нерівномірного розподілу класів. У дослідженні F1-score дозволив об'єктивно оцінити покращення якості персоналізації після інтеграції потокової обробки даних.

5. Latency (затримка обробки) визначається як середній час між моментом надходження події (запиту користувача) та формуванням адаптивного контенту:

$$Latency = t_{response} - t_{request}$$

Ця метрика є критичною для систем реального часу. У контексті персоналізованих сервісів затримка понад 200–300 мс може негативно впливати на користувацький досвід.

Порівняння з базовим підходом. Для перевірки ефективності потокової методології було виконано порівняння з пакетним (batch) підходом, у якому оновлення моделі здійснювалося один раз на 24 години. В обох випадках використовувався однаковий датасет: 1 200 000 подій користувацької взаємодії, 48 000 унікальних користувачів, 12 500 одиниць контенту та 30 днів історичних даних. Дані збиралися у потоковому режимі через ETL-конвеєр, описаний вище, з використанням мікробатчів 1 секунда.

Ключова відмінність полягає у частоті оновлення профілів користувачів: у batch-моделі профіль є статичним протягом доби, а у streaming-моделі використовується мікробатч 1 секунда та інкрементальне оновлення. Таке порівняння дозволяє оцінити вплив реактивності системи на якість персоналізації.

Результати експерименту наведено у таблиці 1.

Таблиця 1

Порівняння показників пакетного та потокового підходів

Підхід	Accuracy	Precision	Recall	F1-score	Latency (мс)
Batch	0.84	0.81	0.79	0.80	390
Streaming	0.90	0.88	0.86	0.87	160

Отримані результати свідчать про зростання показника Accuracy на 6 процентних пунктів порівняно з пакетним підходом, що вказує на загальне підвищення коректності персоналізації. Одночасно спостерігається покращення значень Precision та Recall, що означає як зменшення кількості нерелевантних рекомендацій, так і більш повне врахування потенційно цікавого для користувача контенту. Це, своєю чергою, зумовило збільшення інтегрального показника F1-score, який відображає збалансованість між точністю та повнотою. Крім того, зафіксовано зменшення показника Latency більш ніж удвічі, що свідчить про суттєве скорочення часу формування персоналізованої відповіді. Виявлені покращення

пояснюються оперативним врахуванням нових поведінкових сигналів користувачів та інкрементальним оновленням профілів у потоковому режимі обробки даних.

Значення Accuracy, Precision, Recall та F1-score (рис. 1) є системно вищими порівняно з пакетним підходом, що свідчить про більш повне та коректне врахування поведінкових характеристик користувачів. Очевидно, що покращення не є локальним для окремої метрики, а має комплексний характер, охоплюючи як точність рекомендацій, так і повноту виявлення релевантного контенту.

Використання потокового ETL-конвеєра дозволило суттєво скоротити час формування персоналізованої відповіді (рис. 2). Зменшення Latency більш ніж удвічі порівняно з пакетною моделлю пояснюється відсутністю тривалих циклів накопичення та періодичного перерахунку профілів. Обробка подій мікробатчами з інтервалом 1 секунда забезпечила майже миттєве оновлення моделі та адаптацію контенту до поточної активності користувача, що є критично важливим для систем реального часу.

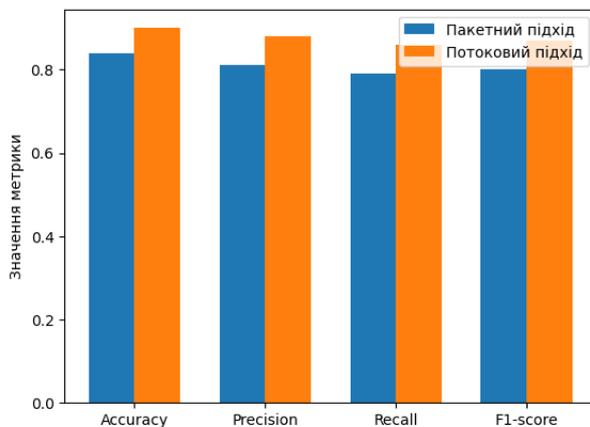


Рис. 1. Порівняння Accuracy, Precision, Recall та F1-score

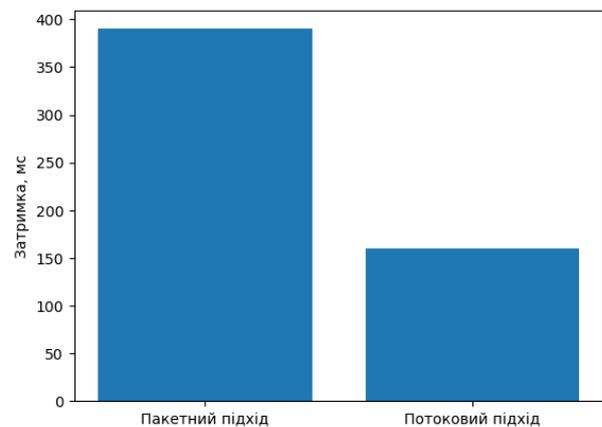


Рис. 2. Порівняння затримки (Latency)

Рисунок 3 демонструє меншу дисперсію значень Accuracy у потоковій конфігурації за результатами 10 незалежних запусків експерименту. Вужчий інтерквартильний розмах та відсутність значних відхилень свідчать про вищу стабільність і передбачуваність роботи моделі. Це означає, що отримані покращення не є випадковими або залежними від конкретного розбиття вибірки, а мають системний характер, що додатково підтверджує надійність запропонованої методології.

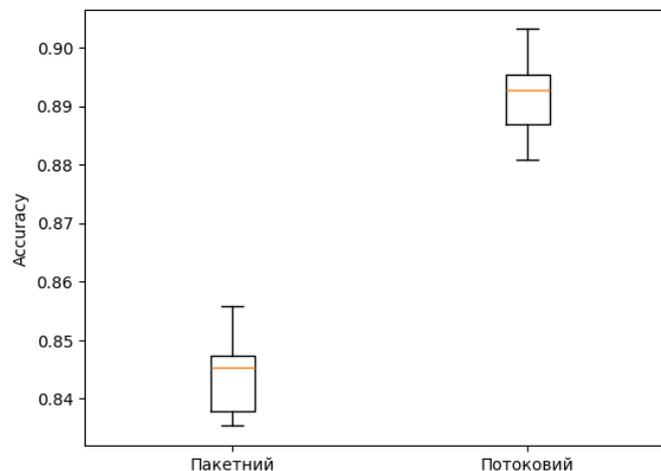


Рис. 3. Розподіл Accuracy (10 запусків)

Таким чином, експериментально підтверджено, що перехід до потокової архітектури забезпечує підвищення якості персоналізації, що проявляється у зростанні показників Accuracy, Precision, Recall та F1-score. Одночасно спостерігається суттєве зменшення значення Latency, що свідчить про скорочення часу формування персоналізованої відповіді та покращення реактивності системи. Проведена статистична перевірка підтвердила значущість отриманих відмінностей, що дозволяє говорити про обґрунтовану перевагу потокового підходу над пакетним. Крім того, результати багаторазових запусків експерименту демонструють меншу варіативність показників, що вказує на підвищення стабільності та передбачуваності функціонування системи.

Висновки. У результаті проведеного дослідження було розроблено та експериментально перевірено методологію побудови системи адаптивної генерації контенту на основі потокової обробки даних. Запропонований підхід інтегрує ETL-конвеєр з мікробатчем 1 секунда, інкрементальне оновлення моделей персоналізації та механізми оперативного формування рекомендацій у режимі, наближеному до реального часу. На відміну від традиційного пакетного підходу з періодичним оновленням профілів користувачів, розроблена архітектура забезпечує безперервну актуалізацію поведінкових характеристик і мінімізує ефект застарілих даних.

Експериментальна перевірка здійснювалася на датасеті, що містив 1 200 000 подій взаємодії 48 000 користувачів з 12 500 одиницями контенту протягом 30 днів. Отримані результати продемонстрували зростання показників Accuracy, Precision, Recall та F1-score порівняно з пакетною моделлю, що свідчить про підвищення якості персоналізації. Одночасно зафіксовано суттєве зменшення Latency, що підтверджує підвищення реактивності системи та її придатність до використання у середовищах реального часу. Проведений статистичний аналіз підтвердив значущість отриманих відмінностей, а результати багаторазових запусків експерименту засвідчили стабільність та відтворюваність показників.

Таким чином, запропонована методологія доводить ефективність використання потокової архітектури для задач персоналізованої генерації контенту. Її впровадження дозволяє одночасно підвищити якість рекомендацій та скоротити час формування відповіді без погіршення стабільності системи.

Перспективи подальших досліджень пов'язані з розширенням експериментальної бази за рахунок збільшення тривалості спостереження та різноманітності поведінкових сценаріїв, інтеграцією більш складних моделей машинного навчання (зокрема глибинних нейронних мереж), а також дослідженням механізмів автоматичної адаптації гіперпараметрів у поточковому середовищі. Окремим напрямом може стати оптимізація використання обчислювальних ресурсів та аналіз впливу масштабування інфраструктури на якісні показники персоналізації.

References

1. Zhang, X., Wei, X., Hu, W. et al. (2025). A Survey on Personalized Content Synthesis with Diffusion Models. *Machine Intelligence Research*, 22, 817–848. DOI: <https://doi.org/10.1007/s11633-025-1563-3>.
2. Yao, Y., & González Vélez, H. (2025). AI Powered System to Facilitate Personalized Adaptive Learning in Digital Transformation. *Applied Sciences*, 15(9), 4989. DOI: <https://doi.org/10.3390/app15094989>.
3. Tiwari, A. (2024). Custom AI Models Tailored to Business Specific Content Needs. *Jurnal Komputer*,

Література

1. Zhang X., Wei X., Hu W. et al. A Survey on Personalized Content Synthesis with Diffusion Models. *Machine Intelligence Research*. 2025. Vol. 22. P. 817–848. DOI: <https://doi.org/10.1007/s11633-025-1563-3>.
2. Yao Y., González Vélez H. AI Powered System to Facilitate Personalized Adaptive Learning in Digital Transformation. *Applied Sciences*. 2025. No. 15(9). Art. 4989. DOI: <https://doi.org/10.3390/app15094989>.
3. Tiwari A. Custom AI Models Tailored to Business Specific Content Needs. *Jurnal Komputer, Informatika*

- Informasi dan Teknologi*, 4(2), 2457. DOI: <https://doi.org/10.53697/jkomitek.v4i2.2457>.
4. Korostin, O. (2025). Optimising Machine Learning Integration in Real Time Text Analytics Platforms: Technical Approaches and Performance Criteria. *Computer Integrated Technologies: Education, Science, Production*, 58. DOI: <https://doi.org/10.36910/6775-2524-0560-2025-58-05>.
5. Adhwaru, H. (2025). Real-Time AI: Building Intelligent Stream Processing Pipelines. *Int. Journal of Scientific Research in Computer Science, Engineering and Information Technology*. DOI: <https://doi.org/10.32628/CSEIT251112313>.
6. Wasilewski, A., Chawla, Y., & Pralat, E. (2025). Enhanced E-Commerce Personalization Through AI Powered Content Generation Tools. *IEEE Access*. DOI: <https://doi.org/10.1109/ACCESS.2025.3550956>.
7. Ismail, A., Sazali, F. H., Jawaddi, S. N. A., & Mutalib, S. (2024). Stream ETL framework for twitter-based sentiment analysis: Leveraging big data technologies. *Expert Systems with Applications*, 261, 125523. DOI: <https://doi.org/10.1016/j.eswa.2024.125523>.
8. Filimonchuk, T. V., Kopailo, Y. R., Partika, S. O., & Sevostyanova, O. M. (2024). Model avtomatyzatsii zboru danykh iz vebresursiv na osnovi heneratyvnoho shtuchnoho intelektu [Automation model for web data collection based on generative AI]. *Information-Control Systems on Railway Transport*, 30(2), 68–79. DOI: <https://doi.org/10.18664/ikszt.v30i2.335414> [in Ukrainian].
9. Chernyshov, H. V., Bredikhin, D. V., Kostenko, O. B., & Bredikhin, V. M. (2025). Intellectual analysis of user behavior in digital services. *Municipal Economy of Cities. Series: Information Technology and Engineering*, 4(192), 9–15. DOI: <https://doi.org/10.33042/3083-6727-2025-4-192-9-15>.
10. Kalashnyk, V. Yu., Melnyk, H. V., & Zlotenko, B. M. (2025). Intelektualizatsiia danykh u rezhymi tsyfrovoho potoku: adaptivni modeli analizu tekstu ta kontekstno-orientovane pryiniattia rishen [Data intelligence in streaming mode: Adaptive text analysis models and context-oriented decision making]. *Science and Technology Today*, 5(46), 1547–1557. DOI: [https://doi.org/10.52058/2786-6025-2025-5\(46\)-1547-1557](https://doi.org/10.52058/2786-6025-2025-5(46)-1547-1557) [in Ukrainian].
11. Halushko, D., Znova, K., & Rolik, O. (2025). Avtomatyzovana systema nalashtuvannya produktiv dlia provaideryv informatsiinykh posluh [Automated product configuration system for information service providers]. *Adaptive Automatic Control Systems*, 1(46), 178–190. DOI: <https://doi.org/10.20535/1560-8956.46.2025.323761> [in Ukrainian].
- dan Teknologi*. 2024. No. 4(2). Art. 2457. DOI: <https://doi.org/10.53697/jkomitek.v4i2.2457>.
4. Korostin O. Optimising Machine Learning Integration in Real Time Text Analytics Platforms: Technical Approaches and Performance Criteria. *Computer Integrated Technologies: Education, Science, Production*. 2025. Vol. 58. DOI: <https://doi.org/10.36910/6775-2524-0560-2025-58-05>.
5. Adhwaru H. Real-Time AI: Building Intelligent Stream Processing Pipelines. *Int. Journal of Scientific Research in Computer Science, Engineering and Information Technology*. 2025. DOI: <https://doi.org/10.32628/CSEIT251112313>.
6. Wasilewski A., Chawla Y., Pralat E. Enhanced E-Commerce Personalization Through AI Powered Content Generation Tools. *IEEE Access*. 2025. DOI: <https://doi.org/10.1109/ACCESS.2025.3550956>.
7. Ismail A., Sazali F. H., Jawaddi S. N. A., Mutalib S. Stream ETL framework for twitter-based sentiment analysis: Leveraging big data technologies. *Expert Systems with Applications*. 2024. Vol. 261. Art. 125523. DOI: <https://doi.org/10.1016/j.eswa.2024.125523>.
8. Філімончук Т. В., Копайло Я. Р., Партика С. О., Севостьянова О. М. Модель автоматизації збору даних із вебресурсів на основі генеративного штучного інтелекту. *Інформаційно-керуючі системи на залізничному транспорті*. 2024. No. 30(2). С. 68–79. DOI: <https://doi.org/10.18664/ikszt.v30i2.335414>.
9. Chernyshov H. V., Bredikhin D. V., Kostenko O. B., Bredikhin V. M. Intellectual analysis of user behavior in digital services. *Municipal Economy of Cities. Series: Information Technology and Engineering*. 2025. No. 4(192). P. 9–15. DOI: <https://doi.org/10.33042/3083-6727-2025-4-192-9-15>.
10. Калашник В. Ю., Мельник Г. В., Злотенко Б. М. Інтелектуалізація даних у режимі цифрового потоку: адаптивні моделі аналізу тексту та контекстно-орієнтоване прийняття рішень. *Наука і техніка сьогодні*. 2025. No. 5(46). С. 1547–1557. DOI: [https://doi.org/10.52058/2786-6025-2025-5\(46\)-1547-1557](https://doi.org/10.52058/2786-6025-2025-5(46)-1547-1557).
11. Галушко Д., Знова К., Ролік О. Автоматизована система налаштування продуктів для провайдерів інформаційних послуг. *Адаптивні системи автоматичного управління*. 2025. No. 1(46). С. 178–190. DOI: <https://doi.org/10.20535/1560-8956.46.2025.323761>.

12. Seenivasan, D. (2023). Real-Time Data Processing with Streaming ETL. *International Journal of Science and Research (IJSR)*, 12(11), 2185–2192. DOI: <https://dx.doi.org/10.21275/SR24619000026>. URL: <https://www.ijsr.net/getabstract.php?paperid=SR24619000026>.
13. Zavgorodnii, V. V., Zavgorodnia, H. A., Valiavska, N. O., Adamenko, V. S., Dorohovtsev, Ye. V., & Nesmachnyi, P. V. (2022). Metod avtomatychnoi heneratsii kontentu na osnovi protsedurnykh alhorytmiv [Method for automatic content generation based on procedural algorithms]. *Scientific Notes of V.I. Vernadsky Taurida National University. Technical Sciences Series*, 33(72), 1, 91–96. DOI: <https://doi.org/10.32838/2663-5941/2022.1/15> [in Ukrainian].
14. Zavgorodnia, H. A., & Zavgorodnii, V. V. (2025). Vykorystannia alhorytmiv mashynnoho navchannia dlia dynamichnoi adaptatsii skladnosti komp'uternykh ihor [Using machine learning algorithms for dynamic adaptation of computer game complexity]. *Tavriyskyi Scientific Bulletin. Technical Sciences Series*, 1(5), 156–163. DOI: <https://doi.org/10.32782/tnv-tech.2025.5.1.16> [in Ukrainian].
15. Zavgorodnia, H. A., & Zavgorodnii, V. V. (2025). Rozrobka masshtabovanoi rozpodilenoї arkhitektury dlia masovykh bahatokorystuvatskykh onlain-system [Development of a scalable distributed architecture for massive multiplayer online systems]. *Bulletin of Kherson National Technical University*, 4(95)-3, 99–106. DOI: <https://doi.org/10.35546/kntu2078-4481.2025.4.3.11> [in Ukrainian].
16. Zavgorodnia, H. A., & Zavgorodnii, V. V. (2025). Modeliuvannia povedinky hravtsia cherez neiromerezhevi ahenty [Modeling player behavior through neural network agents]. *Scientific Notes of V.I. Vernadsky Taurida National University. Technical Sciences Series*, 36(75), 5-2, 141–145, 2025. DOI: <https://doi.org/10.32782/2663-5941/2025.6.2/20> [in Ukrainian].
12. Seenivasan D. Real-Time Data Processing with Streaming ETL. *International Journal of Science and Research (IJSR)*. 2023. No. 12(11). P. 2185–2192. DOI: <https://dx.doi.org/10.21275/SR24619000026>. URL: <https://www.ijsr.net/getabstract.php?paperid=SR24619000026>.
13. Завгородній В. В., Завгородня Г. А., Валявська Н. О., Адаменко В. С., Дороговцев Є. В., Несмачний П. В. Метод автоматичної генерації контенту на основі процедурних алгоритмів. *Вчені записки ТНУ ім. Вернадського. Серія: Технічні науки*. 2022. Т. 33(72), № 1. С. 91–96. DOI: <https://doi.org/10.32838/2663-5941/2022.1/15>.
14. Завгородня Г. А., Завгородній В. В. Використання алгоритмів машинного навчання для динамічної адаптації складності комп'ютерних ігор. *Таврійський науковий вісник. Серія: Технічні науки*. 2025. No. 1(5). С. 156–163. DOI: <https://doi.org/10.32782/tnv-tech.2025.5.1.16>.
15. Завгородня Г. А., Завгородній В. В. Розробка масштабованої розподіленої архітектури для масових багатокористувачьких онлайн-систем. *Вісник Херсонського національного технічного університету*. 2025. No. 4(95), Ч. 3. С. 99–106. DOI: <https://doi.org/10.35546/kntu2078-4481.2025.4.3.11>.
16. Завгородня Г. А., Завгородній В. В. Моделювання поведінки гравця через нейромережеві агенти. *Вчені записки ТНУ ім. Вернадського. Серія: Технічні науки*. 2025. Т. 36(75), № 5, Ч. 2. С. 141–145, 2025. DOI: <https://doi.org/10.32782/2663-5941/2025.6.2/20>.

ZAVHORODNIA HANNA

Candidate of Technical Sciences, Associate Professor,
Department of Computer Engineering,
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute", Ukraine
<https://orcid.org/0000-0001-8523-1761>
Scopus Author ID: 57216155533
Researcher ID: PLR-2465-2026
E-mail: annzavgorodnya@gmail.com

ZAVHORODNII VALERII

Doctor of Technical Sciences, Professor,
Department of Computer Engineering,
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute", Ukraine
<https://orcid.org/0000-0002-8347-7183>
Scopus Author ID: 57184425000
Researcher ID: P-5232-2018
E-mail: zavgorodniivalerii@gmail.com

TKACHENKO OLHA

Doctor of Technical Sciences, Professor,
Department of Software Systems and Technologies,
Taras Shevchenko National University of Kyiv, Ukraine
<https://orcid.org/0000-0001-7983-9033>
Scopus Author ID: 59155724200
E-mail: okar@ukr.net

SAVCHENKO ANDRII

Candidate of Technical Sciences,
Department of Information and
Communication Technologies
Higher Education Institution «Academician Yuri Bugay
International Science and Technical University»
<https://orcid.org/0000-0002-8314-6034>
E-mail: an.savchenko@istu.edu.ua

**Hanna ZAVHORODNIA¹, Valerii ZAVHORODNI¹,
Olha TKACHENKO², Andrii SAVCHENKO³**

¹ National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Ukraine

² Taras Shevchenko National University of Kyiv, Ukraine

³ Higher education Institution «Academician Yuri Bugay International Science
and Technical University», Kyiv, Ukraine

METHODOLOGY FOR DEVELOPING AN ADAPTIVE CONTENT GENERATION SYSTEM

Purpose. The aim of this article is to develop a methodology for creating an adaptive content generation system that ensures a high level of personalization, scalability, and integration of modern machine learning algorithms. Special attention is given to staged data processing through ETL pipelines, real-time stream processing with 1-second micro-batches, and the integration of ML models for dynamic content adaptation based on user behavior.

Methodology. This study proposes a combined approach that unites ETL processes for data preparation and normalization, stream processing to ensure real-time performance and minimal latency, and the integration of ML models with incremental updates for content prediction and personalization. The methodology includes a formal system representation through functional blocks, mathematical data processing models, and modern frameworks for scalable computation, enabling flexible management of large event streams and diverse user scenarios.

Findings. Experimental evaluation was conducted on a dataset containing 1,200,000 user interaction events, 48,000 unique users, and 12,500 content items over a 30-day period. Comparison between streaming and batch approaches demonstrated a 6-percentage-point increase in Accuracy, improvements in Precision, Recall, and F1-score, as well as more than a twofold reduction in Latency. Visual analysis and statistical validation confirmed the stability and predictability of the streaming architecture, indicating comprehensive improvements in both personalization quality and system performance.

Originality. A formalized model of an adaptive content generation system is proposed, integrating ETL processes, stream processing with micro-batches, and ML models within a unified architecture. Incremental model updating algorithms, dynamic load balancing, and real-time adaptive content adjustment mechanisms have been developed, which have not been described in current literature.

Practical value. The proposed methodology can be applied in recommendation systems, online learning platforms, games, and other digital services where content personalization and adaptability are critical. It reduces data processing costs, enhances user experience, ensures system stability and predictability, and facilitates the integration of modern machine learning algorithms into corporate solutions.

Keywords: adaptive content generation; ETL processes; stream processing; machine learning; personalization; scalable systems.