



ВЕБПЛАТФОРМА ДЛЯ ПЕРСОНАЛІЗОВАНИХ РЕКОМЕНДАЦІЙ ТОВАРІВ НА ОСНОВІ СЕМАНТИЧНОГО АНАЛІЗУ ПОШУКОВИХ ЗАПИТІВ ЗАСОБАМИ ШТУЧНОГО ІНТЕЛЕКТУ

Досліджено проблему незадовільної якості процесу пошуку товарів і відсутності персоналізованих їх рекомендацій в українському сегменті електронної комерції, де провідні агрегатори цін (Hotline.ua, E-Katalog, Price.ua) реалізують переважно лексичне зіставлення назв товарів без семантичного розуміння запитів. Запропоновано архітектуру вебплатформи, що поєднує трансформерну модель Sentence-BERT для семантичного аналізу природномовних запитів з контентною та колаборативною фільтрацією інтересів покупця в єдиній композиційній моделі ранжування товарів, у якій підсумкову релевантність пошуку товару визначають як зважену комбінацію семантичної, контентної та колаборативної компонент. Розроблено композиційну математичну модель оцінювання релевантності пошуку товару у системі персоналізованих їх рекомендацій товару, яка інтегрує результати семантичного аналізу пошукових запитів, контентні характеристики товарів і дані про попередню взаємодію користувача із системою. Особливістю запропонованої моделі є адаптивне налаштування вагових коефіцієнтів окремих складових релевантності пошуку товару залежно від обсягу накопиченої історії взаємодії користувача. Систему реалізовано на тривірневій архітектурі: клієнтський рівень (Next.js/TypeScript), серверний рівень (FastAPI/Python), інтелектуальний рівень (sentence-transformers, implicit). Для зберігання 384-вимірних векторних подань товарів використано базу даних PostgreSQL 16 з розширенням pgvector, що забезпечує наближений пошук методом найближчих сусідів. Реалізовано механізм пояснюваності рекомендацій товару через декомпозицію підсумкового рейтингу аналогічних товарів за внесками семантичної, контентної та колаборативної компонент із текстовими поясненнями у користувацькому інтерфейсі. Реалізовано конекторну архітектуру агрегації даних із зовнішніх джерел з паралельним асинхронним їх опитуванням і обробленням відмов у режимі graceful degradation, що забезпечує відкритість системи щодо нових джерел даних без зміни її ядра. Підтверджено offline-оцінюванням на напівсинтетичному наборі даних перевагу адаптивного зважування компонент ранжування товарів над фіксованим їх вагомостями (приріст за метрикою NDCG@10 +0,095 при $|H|=0$ та +0,059 при $|H|=50$) і відсутність стрибків якості ранжування товарів, властивих дискретному перемиццю алгоритмів. Окреслено перспективи подальших етапів дослідження: online A/B-оцінювання якості ранжування товарів на реальному трафіку, басівська оптимізація гіперпараметрів моделі.

Ключові слова: векторні подання пошукових запитів; адаптивні вагові коефіцієнти компонент; холодний старт рекомендаційної системи; колаборативна фільтрація вподобань користувачів; робота з векторними даними.

Вступ / Introduction

Електронна комерція є однією з найдинамічніших галузей цифрової економіки розвинених країн. Український ринок електронної комерції впродовж останніх років демонструє стійке двозначне зростання: у 2020 р. його обсяг зріс на 48 %, а у 2021 р. – на 27 % за світового темпу близько 15 % [9]. За підсумками 2024 р., українці придбали товарів і послуг онлайн на суму близько 239 млрд грн (на 25 % більше, ніж роком раніше), а кількість активних онлайн-покупців перевищила 11 млн осіб [7]. Рекомендаційні системи стали ключовим чинником доходів провідних онлайн-платформ: за публіч-

ними оцінками [4, 14], модуль Amazon генерує близько 35 % загального доходу компанії, а алгоритм Netflix відповідає за 75–80 % переглядів контенту; методологічні основи цих систем – item-to-item колаборативна фільтрація вподобань користувачів та матрична факторизація "користувач-товар" – описано у працях [13, 16]. Аналізуючи стан галузі [12], підтверджено, що гібридні підходи, які поєднують контентну та колаборативну фільтрацію рекомендацій, домінують у виробничих системах [12, 20].

Попри світову зрілість рекомендаційних технологій, провідні українські агрегатори цін (Hotline.ua, E-Katalog, Price.ua) реалізують переважно лексичне зіставлен-

© Copyright 2026 by the author(s)

Бовтач Мар'ян Іванович, магістрант, кафедра інженерії програмного забезпечення.

Email: magmbov@gmail.com; <https://orcid.org/0009-0001-2863-3892>

Яцишин Світлана Іванівна, канд. техн. наук, доцент, завідувачка кафедри інженерії програмного забезпечення.

Email: yatsyshyn@nltu.edu.ua; <https://orcid.org/0000-0001-5200-4837>

Цитування за ДСТУ: Бовтач М. І., Яцишин С. І. Вебплатформа для персоналізованих рекомендацій товарів на основі семантичного аналізу пошукових запитів засобами штучного інтелекту. Науковий вісник НЛТУ України. 2026, т. 36, № 4. С. 09–16.

Citation APA: Bovtach, M. I., & Yatsyshyn, S. I. (2026). Web platform for personalized product recommendations based on semantic analysis of search queries using artificial intelligence. *Scientific Bulletin of UNFU*, 36(4), 09–16. <https://doi.org/10.36930/40360401>

ня товарів за артикулом або точною їх назвою без семантичного розуміння намірів користувача. Практика електронної комерції свідчить [15], що пошуковий рядок є основною точкою входу користувача на платформу, а незадовільні результати пошуку призводять до втрати значної частини потенційних покупців. Поява архітектури Transformer [22] та двоспрямованої моделі BERT [2], а також її прикладної модифікації Sentence-BERT [18] уможливила перехід від лексичного до семантичного пошуку із підтримкою понад 50 мов, у т.ч. з українською.

Актуальність дослідження зумовлена розривом між сучасним рівнем досконалості рекомендаційних технологій і станом українського онлайн-рітейлу, де пошук товарів досі ґрунтується на лексичному зіставленні товарів без семантичного розуміння запитів і персоналізації рекомендацій. Це знижує релевантність видачі результатів запиту та призводить до втрати покупців, що потребує розроблення й подальшого дослідження семантично орієнтованих персоналізованих рекомендаційних рішень для україномовного середовища.

Об'єкт дослідження – формування персоналізованих рекомендацій товарів у системах електронної комерції на основі аналізу пошукових запитів.

Предмет дослідження – методи, моделі та алгоритми семантичного аналізу пошукових запитів і формування персоналізованих рекомендацій товарів, що дасть змогу підвищити релевантність видачі результатів запиту та якість персоналізації інтересів покупця.

Мета роботи – розробити вебплатформу для персоналізованих рекомендацій товарів на основі семантичного аналізу пошукових запитів засобами штучного інтелекту, яка дасть можливість підвищити релевантність пошуку товару та забезпечити персоналізацію видачі результатів запиту з урахуванням інтересів покупця для українського ринку електронної комерції.

Для досягнення зазначеної мети визначено такі основні завдання дослідження:

- проаналізувати сучасний стан українського ринку електронної комерції та систематизувати світовий досвід провідних рекомендаційних систем, що дасть змогу виявити функціональні обмеження наявних українських агрегаторів цін й обґрунтувати потребу в семантично орієнтованому рішенні;
- дослідити теоретичні основи методів оброблення природної мови, нейромережових векторних подань речень та контентно-колаборативної фільтрації інтересів покупця, що дасть можливість обрати модельний апарат, що відповідає завданню персоналізованих товарних рекомендацій для україномовних запитів;
- розробити трирівневу концептуальну архітектуру вебплатформи з відкритим конекторним рівнем агрегації даних, що забезпечить модульність системи та можливість під'єднання нових магазинів без зміни її ядра;
- спроектувати інформаційне забезпечення рекомендаційної системи на базі реляційної СКБД PostgreSQL 16 з розширенням pgvector, яке уможливить зберігання щільних векторних подань товарів і наближений їх пошук методом найближчих сусідів засобами SQL;
- сформулювати композитну математичну модель ранжування товарів з адаптивними ваговими коефіцієнтами компонент моделі, що дасть змогу єдиним формальним апаратом коректно опрацювати як проблему холодного старту пошукової системи, так і зрілий профіль користувача без дискретного перемикання між алгоритмами;
- реалізувати програмно-алгоритмічне забезпечення серверної та клієнтської частин пошукової системи з меха-

нізмом пояснюваності рекомендацій товару, що допоможе підвищити прозорість видачі результатів запиту та довіру користувача до результатів роботи системи;

- провести оцінювання запропонованої моделі на наборі даних за метриками NDCG@10 і Precision@10, що дасть можливість емпірично підтвердити перевагу адаптивного зважування компонент ранжування товарів над фіксованими їх вагомостями та дискретним перемиканням між алгоритмами.

Аналіз останніх досліджень та публікацій. Використання штучного інтелекту для семантичного аналізу пошукових запитів дає можливість розуміти саме намір користувача (search intent), а не просто зіставляти ключові слова його намірів. Це значно підвищує точність рекомендацій та конверсію інтернет-магазинів.

У роботі [18] обґрунтовано підхід до побудови щільних векторних подань речень, на основі моделі Sentence-BERT, узгоджених зі смисловою близькістю фраз; багатомовна модель paraphrase-multilingual-MiniLM-L12-v2 підтримує понад 50 мов і забезпечує істотно швидше та точніше семантичне зіставлення коротких текстів. Проте, тут йдеться тільки про семантичне зіставлення запитів без персоналізації інтересів покупця та без механізму поєднання з поведінковою історією користувача.

У роботі [16] побудовано персоналізований рушій рекомендацій товарів для електронної комерції на основі неявних поведінкових сигналів і стратегій пошуку, що покращує ключові бізнес-метрики та частково пом'якшує холодний старт. Однак, запропоноване поєднання компонент є фіксованим і не передбачає неперервної адаптації вагових коефіцієнтів до стадії зрілого профілю користувача.

У роботі [13] запропоновано вдосконалений метод колаборативної сингулярної факторизації (Co-SVD) зі зменшенням розрідженості матриці взаємодій для пом'якшення проблеми холодного старту пошукової системи в електронній комерції, що підвищує точність рекомендацій для користувачів з обмеженою історією. Однак, метод спирається тільки на матрицю взаємодій "користувач-товар" і не використовує семантичного аналізу природномовних запитів.

У роботі [17] запропоновано гібридну рекомендаційну систему, що поєднує колаборативну та контентну фільтрацію для підвищення якості та повноти рекомендацій. Проте, поєднання компонент здійснюється за статичними правилами без урахування обсягу історії взаємодій користувача і без семантичного опрацювання їх запитів.

У роботі [20] пояснюваність видачі рекомендацій розглянуто як один із п'яти ключових викликів галузі, оскільки відсутність прозорого пояснення видачі результатів запиту знижує довіру користувачів. Однак, йдеться про концептуальні виклики у домені музичних рекомендацій без конкретного механізму пояснюваності вибору музикальної композиції для електронної комерції.

У роботі [1] проаналізовано проблему упереджень (bias) у рекомендаційних системах як суміжну із завданням забезпечення пояснюваності вибору. Проте робота не торкається питань адаптивного зважування компонент ранжування товарів і семантичного опрацювання україномовних запитів.

У роботі [24] систематизовано методи глибинного навчання моделей для рекомендаційних систем і виокремлено критичність коректного опрацювання холодно-

го старту пошукової системи та збалансованого поєднання компонент ранжування товарів. Однак, аналіз зосереджено переважно на великих корпоративних платформах та англomовному контенті без єдиного формального апарату для багатомогазинного україномовного середовища.

У роботі [19] сформульовано загальні вимоги до гібридних архітектур та наголошено на важливості явного врахування проблеми холодного старту пошукової системи. Проте робота має концептуальний характер і не пропонує конкретного закону адаптації вагових коефіцієнтів компонент.

Здійснений аналіз останніх досліджень та публікацій свідчить, що наявні роботи розглядають окремі складові завдання – семантичний пошук, колаборативну чи контентну фільтрацію інтересів покупця, пояснюваність вибору – переважно в контексті великих корпоративних платформ та для англomовного контенту. Водночас, жодна з відомих праць не пропонує цілісного рішення для українського ринку, яке поєднувало б семантичний аналіз природномовних україномовних запитів, неперервне адаптивне управління ваговими коефіцієнтами компонент ранжування товарів залежно від стадії профілю користувача та відкрити багатомогазинну архітектуру конекторів. Усе це свідчить про доцільність проведення дослідження з розроблення відповідної вебплатформи.

Матеріали та методи дослідження. У роботі застосовано методи системного аналізу для декомпозиції завдання та формування архітектури системи; методи оброблення природної мови (токенізація, лематизація, виділення ключових сутностей засобами spaCy uk_core_news_sm) для попереднього оброблення пошукових запитів; математичний апарат векторного простору документів (модель TF-IDF та нейромережіві векторні подання пошукових запитів на основі архітектури Transformer) для семантичного порівняння запитів і товарів; методи машинного навчання, зокрема – алгоритми матричної факторизації (англ. *Alternating Least Squares*) та контентної фільтрації інтересів покупця; методи об'єктно-орієнтованого програмування.

Реалізацію системи виконано з використанням таких технологій: серверна частина та ML-модуль – Python (FastAPI 0.110+, SQLAlchemy 2.0, sentence-transformers 2.2+, scikit-learn, implicit, spaCy 3.7+); клієнтська частина – TypeScript (Next.js 14, React 18, Tailwind CSS); сховище даних – PostgreSQL 16 з розширенням pgvector 0.8+; кеш – Redis 7. Векторизацію тексту виконує модель paraphrase-multilingual-MiniLM-L12-v2 розмірністю 384 виміри. Для дедуплікації товарів, отриманих з різних джерел, використано бібліотеку RapidFuzz.

Результати дослідження та їх обговорення / Research results and their discussion

Концептуальна архітектура вебплатформи. Для платформи "Resogo" (умовна робоча назва програмного забезпечення, яка утворена від англ. recommendation + go) запропоновано її трирівневу архітектуру: клієнтський рівень, серверний рівень бізнес-логіки та інтелектуальний рівень оброблення даних. Кожен рівень можна розробляти й замінювати незалежно.

Клієнтський рівень платформи реалізовано як гібрид SSR-сторінок та SPA на Next.js. Серверний рендеринг важливий з двох причин: пришвидшення першого за-

вантаження і забезпечення коректної індексації публічних сторінок. Основні функції: приймання пошукового запиту, відображення результатів і рекомендацій товару, збір неявних сигналів (натискання, час перегляду, додавання до обраного), керування профілем користувача. Взаємодія з сервером – винятково відбувається через REST/JSON.

Серверний рівень (FastAPI/Python) платформи виступає диспетчером запитів: валідує їх через Pydantic, паралельно звертається до зовнішніх джерел через конектори (asyncio.gather() з тайм-аутом 3-5 с), агрегує результати та передає їх інтелектуальному рівню для остаточного ранжування товарів (рис. 1). Інтелектуальний рівень зосереджує всі AI-компоненти: модуль попереднього оброблення текстів (spaCy uk_core_news_sm), модуль векторизації (TF-IDF як baseline та Sentence-BERT як основна модель), модуль композитного ранжування товарів, модуль колаборативної фільтрації інтересів покупця (бібліотека implicit, алгоритм ALS).

Джерела даних оформлено у вигляді уніфікованих конекторів. На етапі мінімально життєздатного продукту (MVP) реалізовано три конектори: SerpApiConnector (пошук через SerpApi: Google Search API), HotlineScraperConnector (парсинг SSR-HTML Hotline.ua), MockConnector (статичний JSON для тестування). Архітектурна модель дає можливість додавати нові джерела (Rozetka, Prom.ua, MakeUp тощо) без зміни ядра. Помилки конекторів обробляються в режимі graceful degradation. Доцільність модульної агрегації сигналів з різних джерел підтверджено у роботі [17].

На рис. 1 наведено діаграму послідовності опрацювання пошукового запиту: клієнт (Next.js) передає запит до API-шлюзу (FastAPI), який звертається до ML-сервісу для векторизації запиту та до конекторів джерел за списком товарів-кандидатів; далі виконується композитне ранжування товарів із формуванням текстових пояснень до рекомендацій.

Логічна модель даних. Як основне сховище даних обрано базу даних PostgreSQL 16 з розширенням pgvector 0.8+, що дає можливість зберігати щільні векторні подання товарів у таблиці бази даних і виконувати ANN-пошук за косинусною подібністю засобами SQL. Розмірність векторів – 384 (відповідає моделі paraphrase-multilingual-MiniLM-L12-v2 і є оптимальним балансом між якістю та обчислювальною вартістю). Для каталогу обсягом до 1 млн товарів обрано індекс IVFFlat з параметром lists = 100; структуру основних таблиць інформаційного сховища наведено в табл. 1.

Композитна математична модель оцінювання релевантності пошуку товару. Композитна (складена, гібридна) математична модель – математичний опис, яка поєднує декілька простіших підмоделей в одну; у цій роботі – модель, яка інтегрує семантичну, контентну та колаборативну компоненти в єдиній зваженій функції релевантності пошуку товару. Завдання формування персоналізованих рекомендацій товарів формалізовано як побудову відображення $R:(q, u) \rightarrow L$, де L – впорядкована за спадом релевантності послідовності товарів для u -го користувача в контексті q -го запиту. В математичній моделі задано множину користувачів $U=\{u_1, \dots, u_N\}$, множину товарів $I=\{i_1, \dots, i_M\}$ та множину взаємодій $H(u)$ u -го користувача.

Семантична подібність між запитом і результатом пошуку. У пошуковій системі q -ий запит користува-

ча та опис i -го товару $d(i)$ проходять конвеєр попереднього оброблення (нормалізацію, лематизацію, токенизацію). Векторні подання речень (sentence embeddings) – числові вектори фіксованої розмірності, у які модель Sentence-BERT кодує зміст природномовного запиту й

опису товару для їх семантичного зіставлення. Далі формуються такі векторні подання [18]:

$$v(q) = E(q), \quad v(i) = E(d(i)), \quad \forall i \in N; \forall q \in Q, \quad (1)$$

де $E(*)$ – функція векторизації (Sentence-BERT, модель paraphrase-multilingual-MiniLM-L12-v2, розмірність 384).

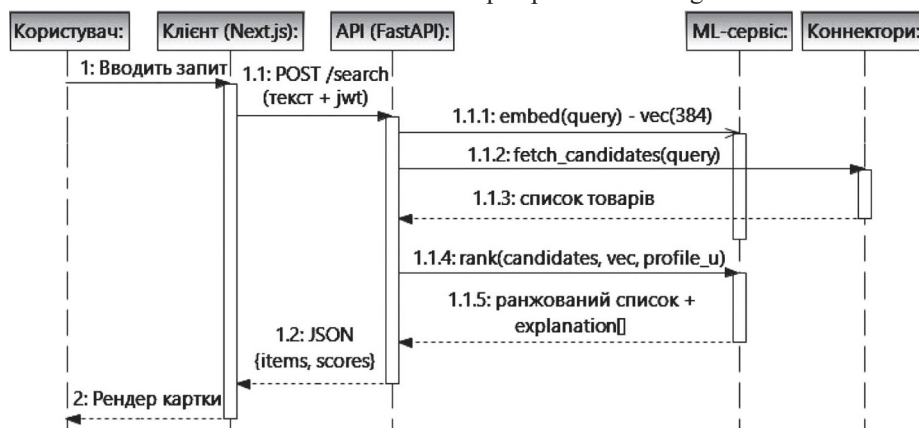


Рис. 1. Діаграма послідовності опрацювання пошукового запиту у вебплатформі / Sequence diagram of search-query processing in the web platform

Табл. 1. Основні таблиці інформаційного сховища системи / Main tables of the system's information storage

Таблиця	Призначення	Ключові поля	Орієнтовний розмір на 1 млн записів
users	Облікові записи	id (UUID), email, password_hash (bcrypt), created_at, prefs	≈200 МБ
items	Каталог товарів	id, title, brand, category, price, description, embedding vector (384)	≈4 ГБ
queries	Журнал запитів	id, user_id, text, query_vector vector (384), timestamp	≈1,5 ГБ
interactions	Взаємодії користувачів	id, user_id, item_id, type, weight, timestamp	≈80 МБ
recommendations	Журнал видач	id, query_id, items[], scores[], weights jsonb	≈600 МБ
sources	Конектори	id, name, type, config jsonb, active	< 1 МБ
als_factors	Кеш ALS-моделі	user_id або item_id, factors vector (64)	≈300 МБ

Семантична подібність між запитом і результатом пошуку обчислюється як косинусна міра [18]:

$$\text{sim}(q, i) = \frac{v(q) \cdot v(i)}{\|v(q)\| \cdot \|v(i)\|}, \quad \forall i \in N; \forall q \in Q, \quad (2)$$

де: $v(q) \cdot v(i)$ – скалярний добуток векторів запиту i -го товару; $\|v(q)\|$, $\|v(i)\|$ – евклідові норми відповідних векторів.

Контентна релевантність пошуку товару. Контентна фільтрація – метод добору товарів за змістовою (семантичною) подібністю запиту до описів характеристик товарів; у математичній моделі відповідає контентній компоненті релевантності. Профіль користувача формується як зважена сума векторів товарів, з якими він взаємодіяв [19]:

$$p(u) = \frac{\sum_{i \in H(u)} w(u, i) \cdot v(i)}{\sum_{i \in H(u)} w(u, i)}, \quad w(u, i) = w_{\text{type}} \cdot \text{Exp}\left(\frac{-\Delta t}{\tau(d)}\right), \quad (3)$$

де: $H(u)$ – історія взаємодій u -го користувача; $w(u, i)$ – вагомість взаємодії як добуток коефіцієнта типу (0,3 – перегляд; 0,7 – перехід у магазин; 1,0 – додавання в обране) та функції експоненційного згасання у часі з часовою сталою $\tau(d) \approx 7$ днів (на відміну від гіперпараметра τ швидкості зростання вагових коефіцієнтів компонент моделі, який вимірюється у кількості взаємодій); $v(i)$ – векторне подання i -го товару.

Контентна релевантність пошуку i -го товару для u -го користувача визначається як косинусна подібність між профілем користувача та вектором товару [19]:

$$\text{rel}_c(u, i) = \cos(p(u), v(i)), \quad \forall i \in N; \forall u \in U, \quad (4)$$

де: $p(u)$ – векторний профіль u -го користувача; $\cos(*, *)$ – функція косинусної подібності між векторами.

Колаборативна релевантність пошуку товару. Колаборативна фільтрація – метод прогнозування вподобань користувача на основі поведінки схожих користувачів (за матрицею "користувач-товар"); у моделі відповідає колаборативній компоненті релевантності. Матрична факторизація – метод лінійної алгебри, що розкладає матрицю взаємодій "користувач-товар" на добуток простіших матриць прихованих факторів; є основою колаборативної компоненти (реалізовано через ALS). Матриця взаємодій R розміру $N \times M$ апроксимується добутком матриць нижчого рангу $R \approx P \cdot Q^T$, де P (розмір $N \times k$) та Q (розмір $M \times k$) визначаються методом ALS для implicit feedback ($k = 64$). Колаборативна релевантність пошуку товару визначається як [6]:

$$\text{rel}_F(u, i) = P_u \cdot Q_i^T, \quad \forall i \in N; \forall u \in U, \quad (5)$$

де: P_u – u -й рядок матриці прихованих факторів користувачів; Q_i^T – транспонований i -й рядок матриці прихованих факторів товарів.

Композитна функція ранжування товарів. Композитне ранжування товарів – упорядкування товарів-кандидатів за підсумковою релевантністю, яку обчислюють як зважену комбінацію семантичної, контентної та колаборативної складових (замість оцінки за одним параметром). Підсумкова релевантність пошуку i -го товару для q -го запиту u -го користувача визначається за такою формулою [6, 18, 19]:

$$\text{rel}(q, u, i) = \alpha \cdot \text{sim}(q, i) + \beta \cdot \text{rel}_c(u, i) + \gamma \cdot \text{rel}_F(u, i), \quad (6)$$

де $\alpha + \beta + \gamma = 1$, $\alpha, \beta, \gamma \geq 0$. Ключовою особливістю математичної моделі ранжування товарів є те, що вагові коефіцієнти не є константами, а адаптуються залежно від обсягу товару $|H(u)|$ за експоненційним законом:

$$\beta(u) = \beta_{\max} \cdot \left(1 - \exp\left(-\frac{|H(u)|}{\tau}\right) \right), \quad (7)$$

$$\gamma(u) = \gamma_{\max} \cdot \left(1 - \exp\left(-\frac{|H(u)|}{\tau}\right) \right), \quad (8)$$

$$\alpha(u) = 1 - \beta(u) - \gamma(u). \quad (9)$$

Стартові значення гіперпараметрів для MVP: $\beta_{\max} = 0,42$; $\gamma_{\max} = 0,28$; $\tau = 8$. Співвідношення $\beta_{\max} > \gamma_{\max}$ відображає вищу надійність індивідуальної історії взаємодій порівняно з колаборативним сигналом за обмеженого обсягу даних, а стала $\tau = 8$ забезпечує помітну активацію персоналізації інтересів покупця вже після 8-10 взаємодій; остаточне калібрування цих значень винесено на етап басівської оптимізації на реальних даних. При $|H(u)| = 0$ маємо $\alpha = 1$, $\beta = \gamma = 0$ (режим "холодного старту пошукової системи" – стан рекомендаційної системи, коли для нового користувача ще немає історії взаємодій, а для нового товару немає відгуків); при $|H(u)| \rightarrow \infty$ маємо $\alpha \rightarrow 0,30$, $\beta \rightarrow 0,42$, $\gamma \rightarrow 0,28$ (зрілий профіль). Плавний перехід відбувається приблизно після 25-30 взаємодій. Таке математичне формулювання дає можливість єдиним апаратом охопити обидва крайніх випадки без дискретного перемикання між алгоритмами (рис. 2).

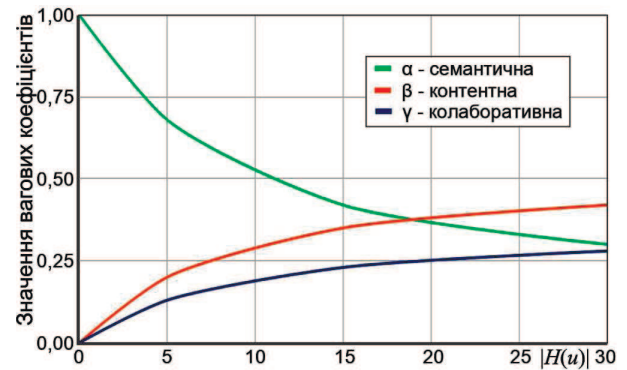


Рис. 2. Динаміка адаптивних вагових коефіцієнтів α , β , γ залежно від обсягу історії взаємодій $|H(u)|$ / Dynamics of adaptive weight coefficients α , β , γ depending on the volume of interaction history $|H(u)|$

Порівняльний аналіз пошукових систем. Запропонований у роботі підхід до рекомендації товарів відрізняється від наявних на ринку класів рішень за типом їх пошуку, рівнем персоналізації інтересів покупця, охопленням джерел, пояснюваністю та способом опрацювання холодного старту пошукової системи (табл. 2). Для проведення порівняльного аналізу найкраще структурувати класи пошукових систем у вигляді порівняльної таблиці.

Табл. 2. Порівняння архітектурних класів пошукових систем у сегменті українського e-commerce / Comparison of architectural classes of systems in the Ukrainian e-commerce segment

Ознака	Агрегатори цін	Маркетплейси	AI-помічники	Recogo
Тип пошуку	Лексичний за артикулом	Лексично-фасетний	Семантичний (LLM)	Семантичний (SBERT) + гібридний
Персоналізація	Відсутня	Один майданчик	Сесійна	Профіль з адаптивними вагомостями
Охоплення джерел	Багатомагазинне	Один майданчик	Каталог не інтегровано	Багатомагазинне через конектори
Пояснюваність	Базова (за ціною)	Не пояснюється	Текстове обґрунтування	Декомпозиція за $\alpha/\beta/\gamma$
Холодний старт	Не актуальний	Популярне в категорії	Не вирішено	Вбудовано в модель ($\alpha \rightarrow 1$)

Це дасть змогу наочно протиставити класичне (наявне на ринку) рішення та запропонований у роботі інноваційний підхід.

Пояснюваність видачі рекомендацій товару реалізовано через декомпозицію рейтингу ранжування товарів за внесками компонент. У відповіді API /search кожен товар супроводжується об'єктом components із трьома значеннями (sim, rel_C, rel_F) та текстове поле explanation. На клієнті це візуалізується у вигляді трьох кольорових індикаторів (semantic / content / collaborative) та пояснювального речення, сформованого за одним з чотирьох шаблонів [19]:

- семантична перевага – "Найкращий смисловий збіг із вашим запитом";
- контентна перевага – "Близько до товарів, які ви розглядали раніше";
- колаборативна перевага – "Популярний серед користувачів зі схожими вподобаннями";
- змішаний випадок – комбінація двох найбільших компонент.

Оцінювання якості ранжування товарів. Для верифікації властивостей моделі проведено offline-оцінювання на напівсинтетичному наборі даних. Сформовано 500 синтетичних профілів користувачів: для кожного профілю з каталогу 200 товарів (384-вимірні одиничні вектори, аналогічні виходу SBERT paraphrase-multilingual-MiniLM-L12-v2) випадково обрано 15 "справді релевантних" товарів; вектор запиту формувалася як цен-

трій цих товарів з гаусовим шумом ($\sigma = 0,3$). Для кожного значення $|H(u)| \in \{0, 2, 5, 8, 10, 15, 20, 30, 50\}$ імітувалась взаємодія з відповідною кількістю товарів із релевантної підмножини; контентний профіль $p(u)$ будувався за цим самим алгоритмом (зважена сума з нормалізацією), а колаборативна компонента апроксимувалася категорійним fallback-алгоритмом (буст за популярністю товарів у відповідній категорії) (табл. 2).

Табл. 3. Метрика NDCG@10 / Precision@10 для трьох стратегій зважування компонент ранжування товарів за різних обсягів історії взаємодій $|H(u)|$ / NDCG@10 / Precision@10 for three weighting strategies at different interaction-history sizes $|H(u)|$

$ H(u) $	Адаптивні вагомості (NDCG@10 / P@10)	Фіксовані вагомості 1/3 (NDCG@10 / P@10)	Дискретне перемикання між алгоритмами (NDCG@10 / P@10)
0	0,285 / 0,259	0,190 / 0,171	0,285 / 0,259
2	0,418 / 0,347	0,417 / 0,273	0,285 / 0,259
5	0,616 / 0,504	0,629 / 0,495	0,636 / 0,496
8	0,715 / 0,611	0,703 / 0,591	0,750 / 0,645
10	0,752 / 0,656	0,732 / 0,629	0,786 / 0,696
15	0,804 / 0,724	0,760 / 0,666	0,815 / 0,737
20	0,847 / 0,780	0,793 / 0,709	0,851 / 0,784
30	0,886 / 0,833	0,828 / 0,755	0,888 / 0,835
50	0,916 / 0,877	0,857 / 0,797	0,916 / 0,877

Порівнювались три стратегії зважування коефіцієнтів компонент ранжування товарів: адаптивні вагомості (поточна модель), фіксовані вагомості $\alpha = \beta = \gamma = 1/3$ та

дискретне перемикання між алгоритмами ($\alpha = 1$ при $|H| < 5$, зрілі вагомості при $|H| \geq 5$). Метрики – NDCG@10 та Precision@10, усереднені за 500 профілями. Результати (табл. 3) підтверджують три ключові властивості адаптивних вагових коефіцієнтів компонент моделі. По-перше, у режимі холодного старту пошукової системи ($|H(u)| = 0$) стратегія з фіксованими вагомостями 1/3 поступається адаптивній вагомості на 0,095 за метрикою NDCG@10, оскільки рівномірна частка колаборативної та контентної компонент додає шум за відсутності профілю. По-друге, дискретне перемикання між алгоритмами залишає якість ранжування товарів на рівні холодного старту пошукової системи при $|H(u)| = 2$ (за метрикою NDCG@10 = 0,285), тоді як адаптивна модель вже досягає 0,418 (приріст + 0,133), плавно інтегруючи ранні сигнали контентної компоненти.

При $|H(u)| = 5$ дискретне перемикання між алгоритмами здійснює стрибок якості ранжування товарів на 0,351 за метрикою NDCG@10 за один крок, що є артефактом порогового алгоритму та може сприйматися користувачем як несподівана зміна поведінки системи. По-третє, у зрілому профілі ($|H(u)| \geq 30$) адаптивна та дискретна стратегії практично зближуються ($\Delta\text{NDCG} \leq 0,002$), тоді як фіксовані вагомості залишаються відстаючими на $0,058^{+0,001}$ за метрикою NDCG@10 через неоптимальний розподіл $\alpha/\beta/\gamma$. Отже, адаптивне зважування компонент ранжування товарів забезпечує монотонне та безперервне зростання якості ранжування товарів від $|H(u)| = 0$ до $|H(u)| \rightarrow \infty$ без видимих стрибків.

Отже, запропонована в роботі композитна модель ранжування товарів з адаптивними ваговими коефіцієнтами вирішує відому суперечність між підходами до проблеми холодного старту пошукової системи. Класичний варіант – дискретне перемикання між алгоритмами за порогоми накопиченої історії – породжує артефакти для користувача: перелік рекомендацій різко змінюється з досягненням порогового значення. Результати offline-оцінювання (див. табл. 3) підтверджують, що, на відміну від дискретного перемикання між алгоритмами, яке демонструє стрибок якості ранжування товарів на 0,351 за метрикою NDCG@10 за один крок, адаптивна модель забезпечує монотонне зростання якості ранжування товарів на (приріст + 0,133 за метрикою NDCG@10 вже при $|H(u)| = 2$).

Обговорення результатів дослідження. На основі семантичного аналізу пошукових запитів для формування персоналізованих рекомендацій товарів можна виокремити кілька ключових аспектів: ефективність розуміння намірів (англ. *Intent Recognition*); вплив на метрики ефективності (KPIs); персоналізація та динамічний контекст; обмеження використання та технічні виклики. Семантичний підхід дає змогу розуміти не просто точні збіги слів, а справжні наміри покупця, контекст запиту та потреби потенційного користувача. Впровадження семантичного аналізу пошукових запитів перетворює пошуковий рядок з інструменту навігації на активного асистента з продажів.

У роботі [12] узагальнено новітні тенденції рекомендаційних систем і відзначено зміщення фокусу до використання гібридних та адаптивних моделей, що узгоджується із запропонованим у цій роботі підходом, проте без явного формалізму неперервної адаптації ваг до обсягу історії взаємодій.

У роботі [5] для пом'якшення проблеми холодного старту пошукової системи запропоновано модель на основі контрастивного навчання з адаптивним злиттям ознак, що динамічно зважає атрибути користувача, метадані товару та контекст запиту. Однак, метод потребує окремого ресурсомісткого етапу навчання, тоді як у запропонованій нашій моделі вагові коефіцієнти задаються детермінованим експоненційним законом без додаткового її навчання, що підтверджено плавним зростанням якості ранжування товарів в табл. 3.

У роботі [11] адаптацію до динамічних сценаріїв холодного старту пошукової системи реалізовано посередньою нейромережею для завдання прогнозування ймовірності кліку (CTR). Однак, такий підхід істотно збільшує складність моделі та вимоги до обсягу навчальних даних, тоді як запропонована єдина композитна функція ранжування товарів досягає схожої мети простішими засобами і зберігає інтерпретованість видачі результатів запиту.

У роботі [15] продемонстровано ефективність семантичного пошуку товарів на промисловому масштабі, що підтверджує доцільність використання щільних векторних подань як основного сигналу релевантності пошуку товару; на відміну від орієнтованого на один майданчик рішення, запропонована в роботі платформа агрегує результати з кількох джерел через конектори.

У роботі [10] запропоновано модель адаптивного навчання гіперпараметрів для глибинного семантичного пошуку, який автоматично балансує внески складників функції втрат. Однак, балансування втрат виконується на етапі навчання моделі, а не як адаптація до обсягу історії конкретного користувача під час роботи пошукової системи; цей напрям корелює з визначеною перспективою – баєсівською оптимізацією гіперпараметрів вагової моделі на реальних даних.

У роботі [8] гібридну рекомендаційну систему для електронної комерції побудовано на неявних оцінках користувачів, що поєднує колаборативну та контентну фільтрації для вирішення проблеми холодного старту пошукової системи. Однак, робота не залучає трансформерних семантичних подань для опрацювання природномовних запитів, тоді як запропонована модель додатково інтегрує семантичну компоненту на основі Sentence-BERT та неперервне адаптивне зважування компонент ранжування товарів.

Використання pgvector замість спеціалізованих ANN-індексів типу Faiss або Annoy спрощує MVP-реалізацію. Обмеженнями поточної реалізації є залежність від доступності зовнішніх API-конекторів і необхідність перевірки на реальному трафіку. Отже, проведене обговорення підтверджує доцільність виконаного дослідження: на відміну від розглянутих робіт, які розв'язують окремі аспекти завдання переважно для англomовного контенту та в межах одного майданчика, запропоноване рішення вперше поєднує неперервне адаптивне зважування компонент ранжування товарів, семантичний аналіз україномовних запитів і відкриту конекторну архітектуру в межах єдиної вебплатформи для українського ринку електронної комерції.

Отже, внаслідок виконаної роботи можна сформулювати такі наукову новизну та практичну значущість результатів дослідження.

Наукова новизна отриманих результатів дослідження – отримала подальший розвиток композитна матема-

тична модель ранжування товарів щодо неперервно адаптивного зважування семантичної, контентної та колаборативної компонент за експоненційним законом залежно від обсягу історії взаємодії користувача, що, на відміну від відомих моделей із дискретним перемиканням між алгоритмами, дає змогу єдиним формальним апаратом коректно опрацьовувати як проблему холодного старту пошукової системи, так і режим зрілого профілю користувача без розривів ранжування товарів.

Практична значущість результатів дослідження – можна використати як основу для побудови вебплатформи персоналізованих рекомендацій товарів для українського ринку електронної комерції, упровадження якої дає змогу підвищити релевантність пошуку товару, скоротити тривалість прийняття рішення про його покупку та забезпечити пояснюваність видачі рекомендацій товару, а також можна застосувати в навчальному процесі під час викладання дисциплін, пов'язаних із розробленням інтелектуальних інформаційних систем.

Висновки / Conclusions

Розроблено вебплатформу "Recogo" для персоналізованих рекомендацій товарів на основі семантичного аналізу пошукових запитів засобами штучного інтелекту, яка дає можливість підвищити релевантність пошуку товару та забезпечити персоналізацію видачі результатів запиту з урахуванням інтересів покупця для українського ринку електронної комерції. За результатами проведеного дослідження можна зробити такі основні висновки.

1. Виявлено системний розрив між світовим рівнем зрілості рекомендаційних технологій та станом українських сервісів: провідні агрегатори цін (Hotline.ua, E-Katalog, Price.ua) реалізують лексичне зіставлення товарів без семантичного розуміння та персоналізації інтересів покупця, що формує незакриту нішу на ринку.
2. Запропоновано трирівневу архітектуру вебплатформи "Recogo" з клієнтським рівнем (Next.js/TypeScript), серверним рівнем (FastAPI/Python) та інтелектуальним рівнем (sentence-transformers, implicit), що забезпечує модульність та можливість незалежної заміни технологій. Архітектура конекторів до зовнішніх джерел даних дає можливість розширювати перелік джерел без зміни ядра системи.
3. Розроблено логічну модель на базі даних PostgreSQL 16 із розширенням pgvector 0.8 для зберігання 384-вимірних векторних подань товарів, що забезпечує ANN-пошук засобами SQL без потреби в окремому векторному сховищі.
4. Розроблено композитну математичну модель ранжування товарів, принциповою особливістю якої є неперервна адаптація вагових коефіцієнтів за експоненційним законом від обсягу історії взаємодії користувача: у режимі холодного старту пошукової системи домінує семантична компонента, а з накопиченням історії зростають внески контентної та колаборативної компонент, що дає змогу єдиним апаратом охопити обидва режими без дискретного перемикання між алгоритмами.
5. Реалізовано механізм пояснюваності рекомендацій товару через декомпозицію фінального рейтингу аналогічних товарів за внесками трьох компонент із візуалізацією кольоровими індикаторами та текстовими поясненнями – підхід, що вирішує проблему "чорної скриньки" довіри до рекомендаційних систем.
6. Проведено offline-оцінювання запропонованої моделі на напівсинтетичному наборі даних, яке підтвердило перевагу адаптивного зважування компонент ранжування товарів над фіксованими їх вагомостями (приріст

NDCG@10 + 0,095 у режимі холодного старту пошукової системи та + 0,059 у зрілому профілі) і відсутність стрибків якості ранжування товарів, властивих дискретному перемиканню алгоритмів, чим емпірично обґрунтовано доцільність запропонованого підходу.

Сукупно отримані результати підтверджують досягнення поставленої мети та практичну придатність розробленої вебплатформи персоналізованих рекомендацій товарів для українського ринку електронної комерції.

Заява про використання інструментів штучного інтелекту. Під час підготовки цього рукопису автори застосовували сервіси штучного інтелекту та програмне забезпечення для оброблення природної мови (ChatGPT, Grammarly та Copilot) винятково задля граматичного редагування, коригування орфографії та стилістичної адаптації складних синтаксичних конструкцій. Автори здійснювали критичний аналіз і подальше редагування згенерованого тексту та несуть повну відповідальність за достовірність, науковий зміст і остаточну редакцію публікації.

References

1. Chen, J., Dong, H., Wang, X., Feng, F., Wang, M., & He, X. (2023). Bias and debias in recommender system: A survey and future directions. *ACM Transactions on Information Systems*, 41(3), 1–39. <https://doi.org/10.1145/3564284>
2. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of NAACL-HLT 2019*, 4171–4186. <https://doi.org/10.18653/v1/N19-1423>
3. Fenii, N. S., & Hrytsiuk, Y. I. (2020). Automation of the process of classification of text news from internet sites by neural network methods. *Scientific Bulletin of UNFU*, 30(4), 123–133. <https://doi.org/10.36930/40300421>
4. Gomez-Urbe, C. A., & Hunt, N. (2015). The Netflix recommender system: Algorithms, business value, and innovation. *ACM Transactions on Management Information Systems*, 6(4), 1–19. <https://doi.org/10.1145/2843948>
5. Hu, J., An, T., Yu, Z., Du, J., & Luo, Y. (2025). Contrastive learning for cold start recommendation with adaptive feature fusion. *arXiv preprint arXiv:2502.03664*. <https://doi.org/10.48550/arXiv.2502.03664>
6. Hu, Y., Koren, Y., & Volinsky, C. (2008). Collaborative filtering for implicit feedback datasets. *Proceedings of the 8th IEEE International Conference on Data Mining (ICDM 2008)*, 263–272. <https://doi.org/10.1109/ICDM.2008.22>
7. Kasumov, T. (2024). Marketplaces in Ukraine in the conditions of digital metamorphoses: Trends, challenges and strategic guidelines. *Journal of Strategic Economic Research*, 25(6), 64–78. <https://doi.org/10.30857/2786-5398.2024.6.7>
8. Keikhosrokiani, P., & Fye, G. M. (2024). A hybrid recommender system for health supplement e-commerce based on customer data implicit ratings. *Multimedia Tools and Applications*, 83(15), 45315–45344. <https://doi.org/10.1007/s11042-023-17321-6>
9. Kolupaieva, I. V., & Bezsonov, S. (2023). The state and directions of development of e-commerce in Ukraine. *Problems of the Economy*, 2(56), 74–80. <https://doi.org/10.32983/2222-0712-2023-2-74-80>
10. Li, M., Yuan, C., Wang, H., Wang, P., Zhuo, J., Wang, B., & Xu, S. (2023). Adaptive hyper-parameter learning for deep semantic retrieval. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Industry Track*, 775–782. <https://doi.org/10.18653/v1/2023.emnlp-industry.72>
11. Li, W., Zhou, J., Luo, C., Tang, C., Zhang, K., & Zhao, S. (2024). Scene-wise adaptive network for dynamic cold-start scenes optimization in CTR prediction. In *Proceedings of the 18th ACM Conference on Recommender Systems (RecSys 24)*, 370–379. <https://doi.org/10.1145/3640457.3688115>
12. Li, Y., Liu, K., Satapathy, R., Wang, S., & Cambria, E. (2024). Recent developments in recommender systems: A survey. *IEEE*

- Computational Intelligence Magazine*, 19(2), 78–95. <https://doi.org/10.1109/MCI.2024.3363984>
13. Loukili, M., & Messaoudi, F. (2025). Enhancing cold-start recommendations with innovative Co-SVD: A sparsity reduction approach. *Statistics, Optimization & Information Computing*, 13(1), 396–408. <https://doi.org/10.19139/soic-2310-5070-2048>
 14. MacKenzie, I., Meyer, C., & Noble, S. (2013). How retailers can keep up with consumers. *McKinsey & Company*. URL: <https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers>
 15. Magnani, A., Liu, F., Chaidaroon, S., Yadav, S., Suram, P. R., Puthenputhussery, A., & Liao, C. (2022). Semantic retrieval at Walmart. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD 22)*, 3495–3503. <https://doi.org/10.1145/3534678.3539164>
 16. Nguyen, D.-N., Nguyen, V.-H., Trinh, T., Ho, T., & Le, H.-S. (2024). A personalized product recommendation model in e-commerce based on retrieval strategy. *Journal of Open Innovation: Technology, Market, and Complexity*, 10(2), article ID 100303. <https://doi.org/10.1016/j.joitmc.2024.100303>
 17. Parthasarathy, G., & Sathiyadevi, S. (2023). Hybrid recommendation system based on collaborative and content-based filtering. *Cybernetics and Systems*, 54(4), 432–453. <https://doi.org/10.1080/01969722.2022.2062544>
 18. Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence embeddings using Siamese BERT-networks. *Proceedings of EMNLP-IJCNLP* 2019, 3982–3992. <https://doi.org/10.18653/v1/D19-1410>
 19. Ricci, F., Rokach, L., & Shapira, B. (2015). Recommender systems: Introduction and challenges. In F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender Systems Handbook* (2nd ed., pp. 1–34). Springer. https://doi.org/10.1007/978-1-4899-7637-6_1
 20. Schedl, M., Zamani, H., Chen, C.-W., Deldjoo, Y., & Elahi, M. (2018). Current challenges and visions in music recommender systems research. *International Journal of Multimedia Information Retrieval*, 7(2), 95–116. <https://doi.org/10.1007/s13735-018-0154-2>
 21. Torskyi, O. I., & Hrytsiuk, Y. I. (2025). Application of machine learning to enhance the efficiency of automated software testing. *Scientific Bulletin of UNFU*, 35(4), 142–149. <https://doi.org/10.36930/40350416>
 22. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008. <https://doi.org/10.48550/arXiv.1706.03762>
 23. Veseliak, V. V., & Hrytsiuk, Y. I. (2024). Machine learning methods in epidemiological research. *Scientific Bulletin of UNFU*, 34(4), 59–67. <https://doi.org/10.36930/40340408>
 24. Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys*, 52(1), 1–38. <https://doi.org/10.1145/3285029>

M. I. Bovtch, S. I. Yatsyshyn

Ukrainian National Forestry University, Lviv, Ukraine

WEB PLATFORM FOR PERSONALIZED PRODUCT RECOMMENDATIONS BASED ON SEMANTIC ANALYSIS OF SEARCH QUERIES USING ARTIFICIAL INTELLIGENCE

Recommender systems have become a core revenue driver in global e-commerce, yet Ukrainian online retail remains dominated by price aggregators that perform exact lexical matching and offer no meaningful personalization. This study designed and implemented, a web platform that combines transformer-based semantic search with hybrid personalized ranking to bridge this gap for the Ukrainian market. The research applied the multilingual Sentence-BERT model (paraphrase-multilingual-MiniLM-L12-v2, 384 dimensions) for dense query and item embedding, Alternating Least Squares on implicit feedback data for collaborative filtering, and exponentially decayed interaction history for content-based user profiling; all three signals are integrated into a single composite ranking function in which the final relevance of an item is a weighted combination of the semantic, content-based, and collaborative components. The scientific novelty of the work lies in the development of a composite mathematical model for assessing product relevance within a personalized recommendation system; this model integrates the results of semantic analysis of search queries, product content characteristics, and data on the user's prior interaction with the system. A distinctive feature of the proposed model is the adaptive adjustment of weight coefficients for individual relevance components based on the volume of the user's accumulated interaction history. The platform stores 384-dimensional item vectors in PostgreSQL 16 database via the pgvector extension with an IVFFlat index. The three-tier architecture – Next.js/TypeScript client, FastAPI/Python server, and a Python intelligence module – is extended by a pluggable connector layer that aggregates results from multiple Ukrainian retailers with graceful degradation. The system exposes an explainability mechanism that decomposes each item's final score into its semantic, content-based, and collaborative contributions, addressing the black-box trust problem. An offline evaluation on a semi-synthetic dataset (500 user profiles, 200 items, 9 history-size levels) shows that adaptive weights outperform fixed equal weights by 0.095 NDCG@10 at cold start and 0.059 at the mature-profile stage, while eliminating the quality jump inherent in discrete algorithm switching. Future work will validate these results on live traffic via online A/B testing.

Keywords: vector representations of search queries; adaptive component weighting coefficients; cold start in recommendation systems; collaborative filtering of user preferences; working with vector data.