

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Національний університет «Запорізька політехніка»



Факультет комп'ютерних наук та технологій
Кафедра «Комп'ютерні системи та мережі»

ВИНИЧЕНКО ДАНИЛ ВІТАЛІЙОВИЧ
Група КНТ-514м

ІНТЕЛЕКТУАЛЬНА АГЕНТО-ОРІЄНТОВАНА
СИСТЕМА ПІДБОРУ ТОВАРІВ

АВТОРЕФЕРАТ

дипломної роботи на здобуття освітньо-кваліфікаційного рівня
«магістр» 123 Комп'ютерна інженерія
освітньої програми Комп'ютерні системи та мережі

2025 р.

Дипломна робота є рукопис.

Робота виконана в національному університеті «Запорізька політехніка», на кафедрі комп'ютерних систем та мереж

Керівник

кандидат технічних наук, доцент
Тягунова Марія Юрївна,
Національний університет «Запорізька
політехніка», доцент кафедри
комп'ютерних систем та мереж

**Офіційний
рецензент:**

Щербаков Володимир Іванович,
Директор НВФ "Бізнес
Комп'ютер Сервіс"

Захист відбудеться "24" грудня 2025 р.

Секретар екзаменаційної комісії, доцент кафедри
комп'ютерних систем та мереж **Т.В. Голуб**

ЗАГАЛЬНА ХАРАКТЕРИСТИКА РОБОТИ

Актуальність теми. У сучасному інтернет-магазині клієнти очікують швидкого та персоналізованого підбору товарів відповідно до своїх потреб та уподобань. Традиційні механізми фільтрації й сортування за фіксованими атрибутами часто не враховують контекст запиту, стиль спілкування або приховані критерії вибору. Це знижує зручність користування вебсистемою та негативно впливає на продажі. Одночасно зростають обсяги товарних каталогів та інформаційне навантаження на користувача, що ускладнює самостійний пошук оптимальних варіантів.

Розвиток технологій штучного інтелекту, зокрема великих мовних моделей (LLM), відкриває нові можливості для побудови рекомендаційних систем та діалогових асистентів. LLM здатні інтерпретувати природномовні запити, поєднувати їх із зовнішнім контекстом та формувати пояснені рекомендації, а RAG-підхід дозволяє підключати зовнішні бази знань і каталоги, зменшуючи кількість хибних або вигаданих відповідей.

Провідні компанії на ринку вже впроваджують подібні рішення: клієнтські асистенти Amazon Rufus, Google Shopping AI, Klarna AI Shopping Assistant, інструменти для продавців Shopify Sidekick та TikTok Seller Assistant, а також сервіс рекомендацій Amazon Personalize. Вони демонструють різні архітектурні підходи – від багатомодельних систем на базі хмарних платформ до класичних рекормендаторів з графами товарів. Водночас аналіз цих рішень виявив низку обмежень – високу вартість та складність інфраструктури, залежність від великих обсягів даних, ризики витоку персональної інформації, упередженості рекомендацій та постійне перезавантаження діалогових вікон при заході на інші сторінки вебсистем.

Агентно-орієнтований підхід дозволяє розподілити складне завдання підбору товарів між спеціалізованими автономними агентами (агент збору контексту, агент роботи з каталогом, агент діалогу, агент оцінювання якості рекомендацій), що підвищує гнучкість, масштабованість та відмовостійкість системи. Поєднання мультиагентної архітектури з LLM та RAG-підходом створює передумови для побудови інтелектуального асистента, який може вести діалог з користувачем, враховувати історію

покупок та інші індивідуальні дані, формуючи персоналізовані добірки товарів та пояснення до них.

Разом з тим, створення власних LLM або складних ML-платформ є економічно недоцільним для малого та середнього бізнесу через значні витрати на апаратне забезпечення й супровід моделей. Тому актуальною є розробка підходів та програмних рішень, які дозволяють орендувати потужності LLM через API, поєднувати їх з локальними модулями пошуку та базами даних і водночас зберігати контроль над критичними бізнес-процесами та користувацькими даними.

Таким чином, актуальною і практично значущою є розробка інтелектуальної агенто-орієнтованої системи підбору товарів для вебсистем електронної комерції, що використовує LLM із RAG-підходом, забезпечує природномовну взаємодію з користувачем, персоналізовані та пояснені рекомендації, а також може бути інтегрована у наявні інтернет-магазини без побудови власної дорогої ML-інфраструктури. Це дає змогу підвищити якість сервісу та конкурентоспроможність онлайн-платформ у умовах швидкого розвитку електронної комерції.

Мета і завдання дослідження. Метою дипломної роботи є підвищення якості та зручності взаємодії з користувачем вебсайту інтернет-магазину за рахунок розробки інтелектуальної агенто-орієнтованої системи підбору товарів на базі великої мовної моделі. Для досягнення цієї мети необхідно:

1. Дослідити особливості агентно-орієнтованих систем та сучасні підходи до використання LLM у задачах підбору товарів;
2. Спроекувати архітектуру інтелектуальної системи підбору;
3. Реалізувати спроектовану вебсистему з ШІ-асистентом;
4. Виконати тестування реалізованої системи та сформулювати рекомендації щодо її застосування.

Об'єктом дослідження є процеси автоматизованого підбору товарів у вебсистемах електронної комерції.

Предмет дослідження – методи, алгоритми та програмні засоби, що забезпечують роботу інтелектуальної агенто-орієнтованої системи підбору товарів на основі LLM, рекомендаційних підходів та вебсервісів.

Методи дослідження. У роботі використано методи аналізу наукових і технічних джерел у галузі рекомендаційних систем,

агентно-орієнтованих підходів та великих мовних моделей. Для побудови архітектури системи використано методи об'єктно-орієнтованого проєктування. Застосовано методи моделювання діалогової взаємодії й промпт-інжинірингу та експериментальне тестування системи в умовах, наближених до роботи реального вебсервісу електронної комерції.

Наукова новизна отриманих результатів:

Наукова новизна полягає в обґрунтуванні комплексного підходу до побудови системи підбору товарів, що поєднує агентно-орієнтовану організацію компонентів, використання LLM з пошуково-посиленням генеруванням (RAG) та сучасні технології розробки у контексті електронної комерції. Запропоновано архітектуру, у якій діалоговий ШІ-асистент, сервіс LLM та вебкомпоненти інтернет-магазину взаємодіють як узгоджена мультиагентна система.

Практичне значення отриманих результатів:

Практичне значення роботи полягає у створенні програмної основи системи ШІ-асистента підбору товарів, яку реалізовано у вигляді вебсервісу з клієнтською частиною на JavaScript/React та серверною частиною на Python/FastAPI з інтеграцією до віддаленої LLM через OpenAI API. Розроблена система забезпечує формування персоналізованих добірок товарів і текстових пояснень, підтримує збереження історії діалогу та може бути інтегрована у вебсайт інтернет-магазину як основа для впровадження персоналізованих сервісів.

Апробація результатів дипломної роботи. Основні ідеї застосування LLM в умовах вебсистем електронної комерції наведено у тезах:

Тягунова М.Ю., Виниченко Д.В. Аналіз асистентів на базі штучного інтелекту для вебсистем. *Тиждень науки-2025. Факультет комп'ютерних наук і технологій*. Тези доп. наук.-практ. конф. (Запоріжжя, 14-18 квіт. 2025). Запоріжжя: НУ «Запорізька політехніка», 2025. С. 69-70. URL: <https://eir.zp.edu.ua/handle/123456789/23295>

Структура та обсяг роботи. Дипломна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел

та додатків. Пояснювальна записка містить 77 сторінок, 19 рисунків та список використаних джерел зі 18 найменувань.

ОСНОВНИЙ ЗМІСТ РОБОТИ

Перший розділ присвячено аналізу предметної області автоматизованого підбору товарів у вебсистемах електронної комерції. Розглянуто класичні рекомендаційні підходи (колаборативна, контентна та гібридна фільтрація), особливості агентно-орієнтованих систем і роль великих мовних моделей та RAG-підходу в задачах формування персоналізованих рекомендацій. Окрему увагу приділено огляду сучасних рішень на базі LLM: Amazon Rufus, Google Shopping AI, Shopify Sidekick, Amazon Personalize, Klarna AI Shopping Assistant та TikTok Shop Seller Assistant, для яких проаналізовано архітектуру, сценарії використання, переваги й обмеження з погляду персоналізації, масштабованості та безпеки даних.

У результаті аналізу зроблено висновок, що для швидкої інтеграції інтелектуального асистента в існуючі вебсервіси електронної комерції доцільно обрати підхід, заснований на орендованій LLM з використанням RAG-підходу. Така схема дозволяє поєднати гнучкість генеративної моделі з контролем над даними, уникнути витрат на власну ML-інфраструктуру та врахувати виявлені недоліки існуючих систем (ризики витоку персональних даних, галюцинації моделей, високу ресурсомісткість багатомодельних рішень). Обраний підхід є основою подальшого проектування інтелектуальної агентно-орієнтованої системи підбору товарів.

У другому розділі дипломної роботи детально розглянуто процес проектування інтелектуальної агентно-орієнтованої системи підбору товарів, що інтегрується у вебсистеми електронної комерції. Сформульовано загальні, функціональні та нефункціональні вимоги до системи, які визначають очікувану поведінку ШІ-асистента, обмеження щодо продуктивності, безпеки й зручності використання для кінцевих користувачів.

У частині загальних вимог підкреслено, що система повинна надавати користувацький інтерфейс у вигляді вебдодатку з

діалоговим вікном взаємодії. Користувач повинен мати можливість формувати свої запити природною мовою, а у відповідь отримувати структуровані рекомендації щодо товарів із текстовими поясненнями. Окремою вимогою є наявність інтерфейсу налаштування параметрів ШІ-асистента для керівників бізнесу, що дозволяє адаптувати стиль спілкування, ступінь деталізації відповідей та стратегії підбору товарів без змін програмного коду серверної частини.

Функціональні вимоги фокусуються на сервісно-орієнтованій архітектурі системи та організації обміну даними між компонентами. Система повинна забезпечувати прийом текстових запитів від користувачів, формування структурованих JSON-повідомлень, розширення контексту за RAG-підходом, побудову аргументованого промпту з кількох логічних частин (системна інструкція, користувацький запит, розширений контекст, правила), відправлення його до зовнішньої LLM та подальшу валідацію отриманої відповіді. Додатково важливий функціонал збереження історії діалогу на стороні клієнта, що дозволяє відновлювати контент після перезавантаження сторінки. Обов'язковими є обробка помилок зовнішнього сервісу та інформування користувача про помилки у зручній формі.

Нефункціональні вимоги пов'язані із продуктивністю, безпекою й надійністю системи. Середній час обробки типового запиту в умовах нормального навантаження не повинен перевищувати п'яти секунд. Передбачено використання протоколу HTTPS для передавання даних, зберігання API-ключів окремо від основної системи та обмеження обсягу переданих до LLM персональних даних. Для захисту від зловживань описано механізми обмеження частоти запитів. Сукупність цих вимог визначає рамки, в яких має функціонувати інтелектуальний асистент у реальних умовах вебсервісу.

Важливим є опис архітектури системи. Запропонований варіант побудовано як сервісно-орієнтований вебсервіс із трьома ключовими компонентами – клієнтською вебчастиною, серверною частиною на базі FastAPI та орендованою LLM-системою. На рисунку 1 наведено загальну архітектуру системи та призначення її основних елементів.

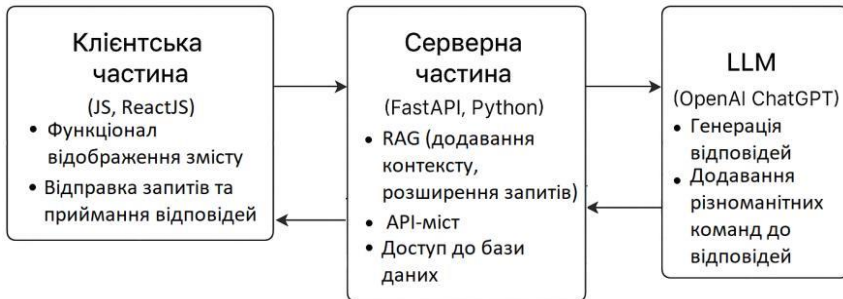


Рисунок 1 – Архітектура системи та її загальне призначення

Клієнтську частину передбачено у вигляді односторінкового вебдодатку з інтерфейсами взаємодії введення користувацьких запитів, відображення відповідей асистента та взаємодію з користувачем. Серверна частина відповідальна за обробку запитів, формування розширеного промπτу, звернення до бази даних та взаємодію з LLM, та за повернення структурованої відповіді на клієнтську частину.

Описано використання RAG-підходу для персоналізації рекомендацій. Після отримання текстового запиту від користувача сервер звертається до локальної бази даних, де зберігаються історія покупок і попередні взаємодії, формує контекст у вигляді набору релевантних записів та інтегрує його в промπτ, який надсилається до LLM. Таким чином, генерація відповіді враховує не тільки поточний запит, а й індивідуальні вподобання користувача. Цей процес частково ємульовано через відсутність повноцінної системи інтернет магазину та її бази даних.

Окремо обґрунтовано вибір зовнішньої LLM ChatGPT як основу інтелектуальної частини системи. Вибір оснований на результатах переддипломної практики, які показали поєднання прийнятної вартості викликів моделі, високої швидкості обробки запитів та відмовостійкості сервісу. Побудова власної мовної моделі потребує значних апаратних та фінансових ресурсів, що робить оренду хмарного сервісу оптимальним варіантом для малого та середнього бізнесу.

Реалізація клієнтської частини на бібліотеці React дозволить використовувати компонентний підхід до побудови інтерфейсу –

чат асистента, панелі налаштувань, блоки з картками товарів та інше виділяються в окремі компоненти з власними станами та життєвим циклом. Такий підхід забезпечить повторне використання коду, полегшить підтримку та надасть змогу оновлювати тільки ті частини інтерфейсу, які змінюються у відповідь на дії користувача чи отримання даних від сервера.

JavaScript, як основна мова для клієнтських вебзастосунків, забезпечує виконання логіки безпосередньо в браузері, що дає змогу реалізувати динамічні сценарії, обробляти події користувача та виконувати асинхронні запити до сервера. React, у свою чергу, суттєво спрощує керування станами інтерфейсу діалогового вікна, дозволяє організувати повторний рендер лише потрібних елементів та підтримує декларативний стиль опису користувацького інтерфейсу. Завдяки цьому, діалог з ІІІ-асистентом запланований без перезавантажень сторінки, а історія повідомлень залишається доступною навіть під час переходу між сторінками системи або зміни окремих компонентів.

Серверна частина спроектована на мові програмування Python із використанням бібліотеки FastAPI, яка забезпечує асинхронну обробку HTTP-запитів та зручні засоби опису REST-інтерфейсів. FastAPI також автоматично генерує документацію для API, що спрощує подальшу інтеграцію інтелектуального агента з іншими сервісами.

Завершальна частина другого розділу присвячена опису діаграми послідовностей, що використовується для моделювання поведінки системи під час обробки користувацького запиту. Діаграма побудовано для демонстрації послідовності обміну повідомленнями між клієнтською частиною, сервером, базою даних та зовнішньою LLM. Вона відображає порядок викликів, активацію компонентів, обробку контексту й повернення результатів, що дає змогу проаналізувати можливі точки збоїв або затримок у роботі системи.

Згідно з діаграмою, яку наведено на рисунку 2, користувач формує текстовий запит у діалоговому інтерфейсі, після чого клієнтська частина перетворює його на HTTP/JSON-повідомлення, додаючи, за потреби, локальний контекст поточної сесії.



Рисунок 2 – Діаграма послідовності обробки користувацького запиту

Сервер приймає запит, звертається до внутрішнього сховища даних для отримання історії покупок і релевантних метаданих, а потім формує розширений промпт. Далі викликається зовнішній LLM-сервіс через OpenAI API, модель генерує відповідь у вигляді структурованого JSON-об'єкта з набором рекомендованих товарів, поясненнями та службовою інформацією. Сервер здійснює валідацію та постобробку результату, після чого повертає його клієнтській частині. Фронтенд оновлює інтерфейс чату, відображаючи як текстову відповідь, так і картки рекомендованих товарів.

Таким чином, у другому розділі сформовано концепцію архітектури інтелектуальної системи підбору товарів, визначено вимоги до неї, обґрунтовано вибір технологій для клієнтської та серверної частин, а також проаналізовано логіку взаємодії між компонентами на основі діаграми послідовностей.

Третій розділ присвячено програмній реалізації інтелектуальної агенто-орієнтованої системи підбору товарів та

перевірці її функціональності. Описано використання орендованої великої мовної моделі через сервіс OpenAI API.

Реалізація програмного прототипу виконана відповідно до сервісно-орієнтованої архітектури, описаної у другому розділі – клієнтська частина, серверна частина на FastAPI та зовнішній LLM-сервіс обмінюються повідомленнями у структурованому форматі. У програмній реалізації використано багатокomпонентний промпт, сформований відповідно до логічної структури, визначеної у другому розділі.

У клієнтській частині реалізовано діалогове вікно ШІ-асистента, механізм збереження історії спілкування у локальному сховищі браузера та відображення рекомендацій у вигляді карток товарів. Передбачено окремий інтерфейс для налаштування промпту, який дозволяє адміністраторам змінювати поведінку асистента без модифікації коду. Серверна частина на Python/FastAPI реалізує точки доступу API для приймання запитів, формування промпту, виклику LLM, валідації та структурування відповіді, а також обробки помилок і роботи демонстраційного режиму(коли немає доступу до LLM).

Описано особливості розгортання системи в локальному середовищі розробника та наведено результати функціонального тестування основних сценаріїв роботи (діалоговий інтерфейс, налаштування промпту, інтеграція з OpenAI API, формування й відображення карток товарів). Критичних відмов не виявлено, зафіксовані недоліки не впливають на працездатність системи й можуть бути усунені в подальшому розвитку системи.

У четвертому розділі узагальнено результати розробки інтелектуальної агенто-орієнтованої системи підбору товарів та оцінюванню її практичної придатності для використання у вебсистемах електронної комерції.

Результатом виконання дипломної роботи є створена система ШІ-асистента підбору товарів, що забезпечує діалогову взаємодію з користувачем та формування персоналізованих рекомендацій на основі великої мовної моделі. Система приймає текстові запити у довільній формі, враховує контекст спілкування та повертає структуровані добірки товарів із поясненнями.

Функціональні можливості системи підтверджено під час тестування. Перевірено роботу діалогового інтерфейсу, збереження

та відновлення історії спілкування, формування відповідей LLM і відображення карток товарів, а також інтерфейс налаштування промпту для зміни поведінки асистента без модифікації коду.

За результатами тестування встановлено, що система стабільно виконує основні сценарії використання, критичних помилок не виявлено. Зафіксовані недоліки не впливають на базову функціональність і не перешкоджають практичному використанню асистента, інтерфейс якого зображено на рисунку 3.

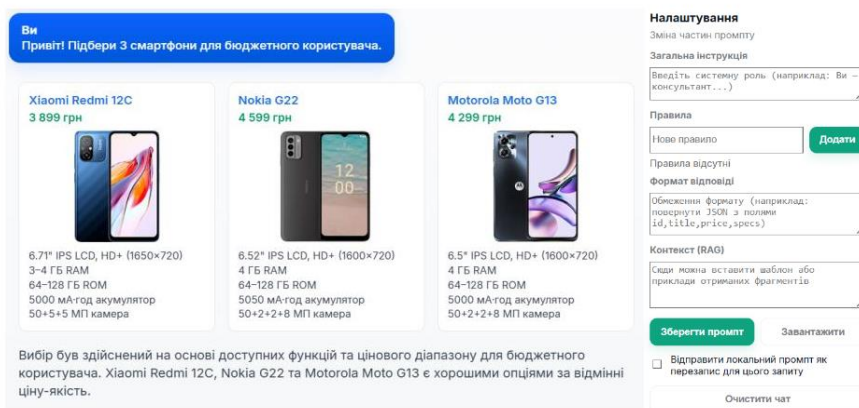


Рисунок 3 – Діалогове вікно ІІІ-асистента підбору товарів та інтерфейс налаштування промпту ІІІ-асистента

Практичне значення роботи полягає у можливості використання розробленої системи як основи для впровадження ІІІ-асистента в інтернет-магазинах з метою покращення якості рекомендацій та зручності взаємодії з користувачами.

ВИСНОВКИ

Результати, отримані в дипломній роботі, є рішенням задачі підвищення якості та зручності підбору товарів у вебсистемах електронної комерції за рахунок використання інтелектуального ІІІ-асистента на основі LLM та підходів, спрямованих на

зменшення типових обмежень існуючих механізмів фільтрації й пошуку.

Отримано такі теоретичні та практичні результати:

Проведено аналіз предметної області та існуючих рішень. Досліджено класичні рекомендаційні підходи, особливості агентно-орієнтованих систем, використання LLM і RAG-підходу в задачах підбору товарів, а також проаналізовано низку актуальних ШІ-асистентів для шопінгу, виявлено їх переваги й обмеження.

Розроблено архітектуру інтелектуального агента підбору товарів. Сформульовано вимоги до системи, запропоновано сервісно-орієнтовану архітектуру з розподілом на клієнтську частину, серверну частину та LLM-сервіс, обґрунтовано використання комбінації сучасних вебтехнологій JavaScript з React та Python з FastAPI та інтеграцією до OpenAI API.

Реалізовано систему з діалоговим інтерфейсом ШІ-асистента, механізмом збереження та відновлення історії діалогу, інтерфейсом налаштувань промпту та режимом повернення структурованих рекомендацій у вигляді карток товарів. Розроблено серверну частину, що формує розширені запити до LLM, виконує валідацію відповідей та забезпечує взаємодію з клієнтом.

Проведено тестування й сформульовано рекомендації щодо застосування. Виконано функціональне тестування основних сценаріїв роботи (діалоговий інтерфейс, історія чату, налаштування промпту, інтеграція з OpenAI API, обробка карток товарів), підтверджено коректність роботи системи та відсутність критичних помилок, запропоновано технічні й бізнес-рекомендації щодо розгортання та подальшого розвитку рішення в інтернет-магазинах.

Таким чином, усі поставлені в роботі завдання виконано, а поставлена мета, розробити прототип інтелектуальної агентно-орієнтованої системи підбору товарів на основі LLM для вебсистем електронної комерції, досягнута.