

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
"КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО"**

Навчально-науковий інститут прикладного системного аналізу

Кафедра штучного інтелекту

"На правах рукопису"

УДК 004.032.26:004.912](043.3)

До захисту допущено:

В. о. завідувачки кафедри

_____ Ірина ДЖИГИРЕЙ

"__" _____ 20__ р.

Магістерська дисертація

на здобуття ступеня магістра

за освітньо-професійною програмою

"Системи і методи штучного інтелекту"

зі спеціальності 122 "Комп'ютерні науки"

**на тему: "Багатодокументне резюмування тексту з використанням
нейромережових методів"**

Виконав:

студент II курсу, групи КІ-41мп

Панасюк Ярослав Ігорович

Науковий керівник:

доцент кафедри ШІ, к.ф.-м.н., доцент

Пишнограєв Іван Олександрович

Рецензент:

доцент кафедри ЕК ФММ, к.ф.-м.н., доцент

Лазаренко Ірина Сергіївна

Засвідчую, що у цій дипломній роботі
немає запозичень з праць інших авторів
без відповідних посилань.

Студент _____

Київ – 2025 року

Національний технічний університет України
"Київський політехнічний інститут імені Ігоря Сікорського"
Навчально-науковий інститут прикладного системного аналізу
Кафедра штучного інтелекту

Рівень вищої освіти – другий (магістерський)

Спеціальність – 122 "Комп'ютерні науки"

Освітньо-професійна програма "Системи і методи штучного інтелекту"

ЗАТВЕРДЖУЮ

В. о. завідувачки кафедри

_____ Ірина ДЖИГИРЕЙ

"29" серпня 2025 р.

ЗАВДАННЯ

на дипломну роботу студенту

Панасюку Ярославу Ігоровичу

1. Тема дисертації "Багатодокументне резюмування тексту з використанням нейромережових методів", науковий керівник дисертації Пишнограєв Іван Олександрович, доцент кафедри ШІ, затверджені наказом по ННІПСА від "06" листопада 2025 р. №4837-с.
2. Термін подання студентом дисертації "13" грудня 2025 року.
3. Об'єкт дослідження: процеси розмітки даних для систем комп'ютерного зору в умовах обмежених маркованих даних.
4. Вихідні дані: система напівавтоматичної розмітки з алгоритмічною передрозміткою та механізмами консенсусу.
5. Перелік завдань, які потрібно розробити:
 - 1) дослідити актуальність обраної теми;
 - 2) здійснити огляд технічної літератури за темою дослідження;

- 3) ознайомитись із існуючими методами багатодокументного резюмування;
 - 4) ознайомитися зі специфікою резюмування відгуків на товари;
 - 5) запропонувати покращення до існуючих рішень;
 - 6) провести порівняння ефективності запропонованих рішень з існуючими;
 - 7) здійснити аналіз комерційного потенціалу та розробити стартап-проект;
 - 8) оформити пояснювальну записку.
6. Перелік ілюстративного матеріалу: презентація.
7. Перелік публікацій:

Заплановано апробацію роботи "Proposed architecture for product review summarisation: a modular pipeline for aspect-based insight extraction and synthesis" автора Panasiuk Yaroslav на IV ВНПК "Системні науки та інформатика" у грудні 2025 р.

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1	Вивчення літератури за темою роботи	02.09.2025-15.09.2025	Виконано
2	Підготовка першого розділу	16.09.2025-20.09.2025	Виконано
3	Підготовка другого розділу	21.09.2025-16.11.2025	Виконано
6	Підготовка третього розділу	21.10.2025-25.11.2025	Виконано
8	Підготовка презентації доповіді	23.10.2025-28.11.2025	Виконано
9	Оформлення дипломної роботи	29.10.2025-13.12.2025	Виконано

Студент

Ярослав Панасюк

Керівник

Іван ПИШНОГРАЄВ

РЕФЕРАТ

Магістерська дисертація: 87 с., 6 рисунків, 16 таблиць, 82 джерела, 1 додаток.

БАГАТОДОКУМЕНТНЕ РЕЗЮМУВАННЯ ВІДГУКІВ, ВЕЛИКІ МОВНІ МОДЕЛІ, АСПЕКТНО-ОРІЄНТОВАНИЙ АНАЛІЗ СЕНТИМЕНТІВ, КІЛЬКІСНЕ АСПЕКТНО-ОРІЄНТОВАНЕ РЕЗЮМУВАННЯ, МЕТРИКИ ОЦІНЮВАННЯ, ФАКТОЛОГІЧНА УЗГОДЖЕНІСТЬ

Об'єкт дослідження – процес автоматизованої обробки та узагальнення текстових відгуків користувачів у системах електронної комерції.

Предмет дослідження – методи та алгоритми аспектно-орієнтованого узагальнення (Aspect-Based Sentiment Summarization), метрики оцінювання фактичної узгодженості згенерованого тексту.

Мета роботи – розробка методології та архітектури системи для трансформації відгуків користувачів у кількісне аспектно-орієнтоване резюме (Quantified Aspect-Based Summary – QABS).

Актуальність роботи полягає у необхідності достовірного й інформативного резюмування надмірної кількості відгуків користувачів.

Наукова новизна полягає у: (1) впровадженні стандарту кількісного аспектно-орієнтованого резюмування; (2) розробці архітектури системи для реалізації кількісного аспектно-орієнтованого резюмування; (3) впровадженні метрик оцінювання фактичної узгодженості згенерованого тексту.

Практична цінність полягає у готовому стандарті та метриках оцінювання резюмування відгуків.

Комерційний потенціал підтверджується високим попитом на інформативні відгуки користувачів як серед покупців, так і продавців товарів та послуг.

Публікація за темою дослідження увійшла до збірника матеріалів IV Всеукраїнської науково-практичної конференції "Системні науки та інформатика".

ABSTRACT

Master's thesis: 86 p., 6 figures, 16 tables, 82 references, 1 appendix.

MULTI-DOCUMENT REVIEW SUMMARIZATION, LARGE LANGUAGE MODELS, ASPECT-BASED SENTIMENT ANALYSIS, QUANTIFIED ASPECT-BASED SUMMARIZATION, EVALUATION METRICS, FACTUAL CONSISTENCY.

The object of research is the process of automated processing and summarization of user text reviews in e-commerce systems.

The subject of research is methods and algorithms of Aspect-Based Sentiment Summarization, and metrics for evaluating the factual consistency of the generated text.

The purpose of the research is the development of a methodology and system architecture for transforming user reviews into a Quantified Aspect-Based Summary (QABS).

The relevance of the work lies in the necessity for reliable and informative summarization of an excessive amount of user reviews.

The scientific novelty consists of: (1) introducing a standard for Quantified Aspect-Based Summarization; (2) developing a system architecture for implementing Quantified Aspect-Based Summarization; (3) introducing metrics for evaluating the factual consistency of reviews summarisation.

The practical value lies in the ready-made standard and evaluation metrics for reviews summarization.

The commercial potential is confirmed by the high demand for informative user reviews among both buyers and sellers of goods and services.

A publication on the research topic was included in the proceedings of the IV All-Ukrainian Scientific and Practical Conference "System Sciences and Informatics".

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ	8
ВСТУП	9
РОЗДІЛ 1 СУЧАСНІ ПІДХОДИ ДО БАГАТОДОКУМЕНТНОГО РЕЗЮМУВАННЯ	11
1.1 Теоретичні основи та еволюція архітектури	11
1.1.1 Перехід від навчання з учителем до навчання зі слабким учителем	11
1.1.2 Архітектури інтеграції багатьох джерел	12
1.1.3 Архітектури з урахуванням зворотного зв'язку та фільтрацією	13
1.2 Агентні системи та системи з доповненим пошуком	14
1.2.1 Інкрементальне оновлення стану узагальнення	15
1.2.2 Генерація з доповненим пошуком (RAG) для відгуків	16
1.2.3 Схеми аргументації на основі аспектів	17
1.3 Метрики та бенчмарки	17
1.3.1 ROUGE та BERTScore	17
1.3.2 Парадигма "LLM-як-суддя"	18
1.3.3 Таксономія та виявлення галюцинацій	19
1.3.4 Калібрування полярності (PoCa)	20
1.4 Порівняльний аналіз підходів та метрик	20
1.5 Набори даних та бенчмарки	23
Висновки до розділу 1	28
РОЗДІЛ 2 ЗАПРОПОНОВАНИЙ ПІДХІД	29
2.1 Визначення цілі: кількісне аспектно-орієнтоване узагальнення (QABS)	29
2.2.1 Критерії якості	30
2.2.2 Структура даних	31

	7
2.3 Порівняння з існуючими реалізаціями резюмування	32
2.4 Архітектура системи	37
2.5 Порівняння з існуючими архітектурами	40
2.6 Методологія оцінювання	41
2.6 Порівняння з існуючими метриками	45
Висновки до розділу 2	46
РОЗДІЛ 3 РОЗРОБКА СТАРТАП-ПРОЕКТУ “ReviewMind AI”	
3.1 Опис ідеї проекту та технологічний аудит	48
3.2 Аналіз ринкових можливостей та конкурентного середовища	56
3.3 Розроблення маркетингової стратегії та бізнес-моделі	61
3.4 Фінансово-економічне обґрунтування проекту	67
ВИСНОВКИ	74
ПЕРЕЛІК ПОСИЛАНЬ	76
ДОДАТОК А ПРИКЛАДИ QAVS	83

ПЕРЕЛІК СКОРОЧЕНЬ

PRS – Product reviews summarisation

SOTA – State of the art

ВСТУП

За останні кілька років напрямок багатодокументного резюмування, а саме резюмування відгуків про товари, зазнав фундаментальних трансформацій, зумовлених стрімким розвитком великих мовних моделей та генеративного штучного інтелекту. В умовах електронної комерції, де користувачі стикаються з тисячами відгуків, що характеризуються високою надлишковістю та суперечливістю, традиційні підходи до обробки тексту втрачають свою ефективність. Напрямок сучасних досліджень змістився від простої генерації тексту у бік створення структурованих висновків – вердиктів та аспектних оцінок, які сприяють швидкому прийняттю рішень. Відтак, задача автоматизованої побудови надійних, фактологічно точних та структурованих резюме набуває критичної значущості для індустрії¹.

Історично проблематика вирішувалася за допомогою екстрактивних методологій, які відбирали найрепрезентативніші речення з корпусу, проте не могли сформулювати цілісні та зв'язні текстові конструкції. Однак представлені трансформерні архітектури (BART, T5, GPT-4) змінили парадигму в бік абстрактивного узагальнення, що дозволила вирішити проблему граматичної зв'язності тексту.

У сучасній науковій літературі виділяють декілька провідних підходів. Фреймворк MEDOS продемонстрував ефективність інтеграції різновидних джерел (відгуки та специфікації) для підвищення фактичної точності. Дослідники Amazon у системі FARSum запропонували враховувати сигнали зворотного зв'язку (актуальність та корисність відгуків) для оптимізації контекстного вікна та генерації структурованих

¹Тут і нижче використано такий інструмент штучного інтелекту як чат-бот з генеративним штучним інтелектом Gemini 3 Pro виключно для корегування та редагування тексту, створеного автором цієї дипломної роботи, на основі автоматизованої перевірки граматики, структури та стилю, що відповідає Політиці використання штучного інтелекту для академічної діяльності в КПП ім. Ігоря Сікорського (протокол №11 Вченої ради КПП ім. Ігоря Сікорського від 11 грудня 2023 р.).

висновків. Іншим важливим напрямом є розробка інкрементальних стратегій, як у фреймворку XL-OPSUMM, що дозволяє обробляти потокові дані без повного перерахунку попередніх даних.

Попри значні досягнення, критичний аналіз існуючих рішень виявляє низку суттєвих недоліків. По-перше, більшість сучасних систем працюють за принципом "чорної скриньки", що призводить до проблеми галюцинацій – генерації тверджень, які суперечать джерелу або містять неіснуючі факти. По-друге, існує "криза оцінювання": традиційні метрики на кшталт ROUGE та BERTScore визнані недостатніми для оцінки якості абстрактивних резюме, оскільки вони не здатні виявити фактологічні помилки та логічні невідповідності. По-третє, існуючі підходи часто не забезпечують верифікованості та квантифікації думок, генеруючи узагальнені фрази замість статистично обґрунтованих тверджень, підкріплених посиланнями на конкретні відгуки. Також залишається відкритою проблема адаптації резюмування під динамічно виникаючі аспекти, яку лише частково вирішують бенчмарки типу OpenAsp.

Враховуючи виявлені прогалини, дослідження було зосереджене на розробці методології, що виходить за межі простого текстового узагальнення в бік "інтелектуального аналізу думок". Було запропоновано підхід до створення Кількісного аспектно-орієнтованого узагальнення (QABS), яке поєднує переваги абстрактивних моделей з жорсткими вимогами до фактологічної узгодженості, верифікованості джерел та використанням новітніх метрик оцінювання на базі LLM (LLM-as-a-Judge).

Проведення дослідження є доцільним та своєчасним, оскільки розробка архітектури, здатної трансформувати неструктурований шум відгуків у верифіковані, квантифіковані та структуровані інсайти, дозволить подолати проблему недовіри до автоматичних систем та забезпечити якісно новий рівень інформаційної підтримки користувачів у системах електронної комерції [1].

РОЗДІЛ 1 СУЧАСНІ ПІДХОДИ ДО БАГАТОДОКУМЕНТНОГО РЕЗЮМУВАННЯ

1.1 Теоретичні основи та еволюція архітектури

Основний виклик в резюмуванні відгуків про товари відрізняє його від загального узагальнення текстів. На відміну від новинної статті, яка зазвичай представляє послідовний виклад фактів, колекція відгуків є багатодокументним вхідним масивом, що характеризується високою надлишковістю, суперечливими суб'єктивними думками та різним рівнем компетентності користувачів. Еволюція архітектур у цій галузі була зумовлена необхідністю розплутати ці складні сигнали.

1.1.1 Перехід від навчання з учителем до навчання зі слабким учителем

У період 2018–2022 років домінував підхід навчання з учителем (supervised learning), де моделі навчалися на наборах даних пар (відгуки, резюме). Проте дефіцит високоякісних, анотованих людиною резюме для мільйонів товарів, доступних онлайн, створив значне вузьке місце. Анотування думок є когнітивно складним і дорогим процесом, що ускладнює масштабування повністю контрольованих підходів на різні категорії продуктів.

Останні дослідження переорієнтувалися на фреймворки без учителя (unsupervised) та зі слабким учителем (weakly-supervised). Наприклад, підхід “Квантований трансформер” (Quantized Transformer, QT) запровадив неконтрольоване екстрактивне узагальнення шляхом перепрофілювання векторно-квантованих варіаційних автоенкодерів (VQ-VAE). Інтерпретуючи

квантований простір як кластери думок, QT зміг ідентифікувати популярні думки без розмічених тренувальних даних. Ця теоретична база заклала основу для сучасних підходів, що використовують створення синтетичних даних (Synthetic Data Creation, SDC) для подолання нестачі розмічених. Використовуючи потужні LLM для генерації псевдомаркованих резюме з відгуків, дослідники тепер можуть донавчати менші й ефективніші моделі на цих синтетичних даних, ефективно дистиллюючи здатність до міркування більшої моделі в архітектуру, придатну до розгортання [4].

1.1.2 Архітектури інтеграції багатьох джерел

Критичним обмеженням ранніх систем PRS була їхня залежність виключно від тексту відгуків користувачів. Відгуки часто містять фактичні неточності, не мають технічного контексту або зосереджуються на нерелевантних проблемах доставки. Для вирішення цієї проблеми SOTA-архітектури еволюціонували у мультиенкодерні фреймворки, що поєднують різновиди джерел даних для створення ґрунтовніших резюме.

Фреймворк MEDOS є яскравим прикладом такої архітектурної еволюції. На відміну від стандартних sequence-to-sequence моделей, які об'єднують усі вхідні дані в єдиний контекстний рядок, MEDOS використовує архітектуру енкодер-декодер із підтримкою гіпермережі та окремими енкодерами для різних типів даних: відгуків користувачів, описів продуктів та пар питання-відповідь [4].

Теоретична перевага такого розділення є суттєвою. Обробляючи опис продукту через виділений енкодер, модель створює сильне представлення специфікацій еталонних даних (наприклад, "цей ноутбук має 8 ГБ оперативної пам'яті"). Одночасно енкодер відгуків фіксує суб'єктивний

настрій користувачів (наприклад, "ноутбук працює повільно"). Декодер потім звертає увагу на обидва представлення, що дозволяє генерувати резюме, які є не лише сентиментально точними, але й фактично узгодженими з офіційними специфікаціями продукту.

MEDOS також запроваджує стратегію "дистиляції мета-знань". Вона передбачає навчання моделі "мета-вчителя", яка фіксує інваріантні для домену знання – особливості та патерни сентиментів, спільні для таких категорій, як електроніка, одяг та товари для дому. Ці мета-знання потім передаються моделі-учню, що значно покращує узагальнення між доменами. Емпіричні результати на наборах даних Amazon, Opossum+ та Flipkart демонструють, що цей мультиджерельний підхід із мета-навчанням забезпечує покращення показників ROUGE-1 F1 на 14,5% порівняно з одноджерельними базовими моделями [4].

1.1.3 Архітектури з урахуванням зворотного зв'язку та фільтрацією

Ще одним значним викликом в обробці потоків відгуків є обробка їхньої зашумленості та визначення актуальності. Якість продукту може змінюватися з часом (наприклад, невдале оновлення прошивки), і не всі відгуки мають однакову якість. Стандартні моделі трансформерів із фіксованими контекстними вікнами часто не можуть пріоритезувати найбільш релевантну інформацію, стикаючись із тисячами відгуків.

Розроблена дослідниками Amazon, архітектура абстрактивного узагальнення з урахуванням зворотного зв'язку (FARSum) вирішує проблему щільності інформації шляхом включення явних сигналів зворотного зв'язку в пайплайн узагальнення. Архітектура ґрунтується на ідеї, що усереднений відгук є менш цінним, ніж корисний відгук..

FARSum реалізує етап суворої фільтрації попередньої обробки на основі двох основних сигналів.

1. Актуальність відгуку: пріоритет надається новішим відгукам для фіксації поточного стану продукту.
2. Зворотний зв'язок клієнтів: використання позначки "Корисно" та зіркових рейтингів для зважування вхідного тексту.

Відфільтрована підмножина високоякісних відгуків подається в абстрактивну модель на базі BART. Важливо, що модель навчена генерувати структуровані резюме, що складаються з трьох окремих сегментів: вердикт, плюси та мінуси. Така структура узгоджується з дослідженнями когнітивної психології щодо того, як споживачі обробляють інформацію: шукають швидкого рішення (вердикту), підкріпленого конкретними доказами (плюси/мінуси) [7].

Інженерним наслідком FARSum є оптимізація контекстного вікна. Фільтруючи саме за корисністю, модель максимізує інформаційну щільність вхідних токенів, зменшуючи ймовірність того, що LLM звертатиме увагу на низькоякісний або нерелевантний текст (наприклад, скарги на доставку). Цей відбір зворотного зв'язку фактично діє як механізм компресії, дозволяючи контекстним вікнам стандартного розміру представляти значно більший корпус думок користувачів [9].

1.2 Агентні системи та системи з доповненим пошуком

Хоча спеціалізовані архітектури, такі як MEDOS і FARSum, представляють вершину донавчених (fine-tuned) систем, ширший індустріальний тренд рухається в бік гнучких пайплайнів, керованих великими мовними моделями (LLM). Ці системи часто використовують

агентні робочі процеси, де LLM діє як керівник, що оркеструє етапи пошуку, планування та узагальнення.

1.2.1 Інкрементальне оновлення стану узагальнення

Для гігантів електронної комерції, таких як Flipkart або Amazon, величезний обсяг вхідних даних робить статичні моделі узагальнення неефективними. Повторний запуск процесу узагальнення на 10 000 відгуків кожного разу, коли надходить новий відгук, є обчислювально неприйнятним.

Інженери Flipkart започаткували стратегію "Інкрементального оновлення", яка розглядає узагальнення як процес зі збереженням стану, а не як одноразове завдання. Основною інновацією є словник аспектів – структура даних, що складається з показників сентиментів споживачів щодо конкретних аспектів (наприклад, батареї, камери, екрану) [10].

Пайплайн працює у структурованому циклі.

1. Ініціалізація: генерується початкове глобальне резюме з пакету відгуків.
2. Інкрементальне оновлення: коли надходить новий пакет відгуків, система не відкидає старе резюме. Натомість вона генерує локальне резюме для нового пакету.
3. Вирішення конфліктів: система використовує словник аспектів як "джерело істини". LLM порівнює існуюче глобальне резюме з новим локальним резюме та статистичними даними у словнику аспектів.
4. Синтез: LLM вирішує конфлікти на основі кількісних доказів та оновлює глобальне резюме [10].

Цей підхід, відомий як "Stateful Summarization", дозволяє системі необмежено масштабуватися з потоковими даними. Він вирішує проблему "катастрофічного забування", притаманну простій заміні старих резюме, та дозволяє уникнути витрат на повторну обробку всієї історії.

1.2.2 Генерація з доповненим пошуком (RAG) для відгуків

Генерація з доповненим пошуком (Retrieval-Augmented Generation, RAG) дозволяє LLM узагальнювати великі масиви відгуків, витягуючи лише найбільш релевантні фрагменти тексту для заповнення контекстного вікна. Однак застосування RAG до відгуків про товари створює унікальні виклики порівняно зі стандартним пошуком документів.

Стандартні пайплайни RAG покладаються на векторні бази даних (використовуючи ембеддинги, такі як text-embedding-3 від OpenAI) для виконання семантичного пошуку. Хоча це ефективно для загальних знань, останні дослідження свідчать, що вектори можуть бути неоптимальними для специфічних запитів щодо відгуків.

Користувач, який шукає "coil whine" (свист котушок) – специфічний високочастотний шум в електроніці, потребує точних збігів. Векторні ембеддинги можуть кластеризувати "coil whine" із загальними термінами, такими як "шум вентилятора" або "гучний", розмиваючи результати пошуку. У таких випадках традиційний пошук за ключовими словами (BM25) часто перевершує векторний пошук, гарантуючи наявність конкретного дефекту у витягнутому контексті [11].

Продуктивні системи все частіше використовують гібридний пошук, поєднуючи точність ключових слів для конкретних функцій або дефектів із семантичним розумінням векторів для загальної тональності, наприклад,

"зручний у використанні". Це гарантує, що контекстне вікно LLM містить як широко релевантну тональність, так і специфічні, важливі деталі [12].

Щоб уникнути надмірності, наприклад, витягування 10 ідентичних відгуків, в яких кажуть "чудова батарея", просунуті пайплайни використовують кластеризацію. Ембеддинги відгуків групуються в кластери, що представляють різні аспекти або полярності тональності. Потім ретривер (retriever) вибирає репрезентативні відгуки з кожного кластера, гарантуючи, що LLM бачить різноманітний зріз думок, а не лише найчисленніші [13]. Ця техніка є важливою для Покриття (Coverage) – забезпечення включення думок меншості (наприклад, 5% користувачів, які повідомляють про небезпеку) у контекст резюме.

1.2.3 Схеми аргументації на основі аспектів

Високотехнологічним підходом, що з'явився у 2024 році, є інтеграція схем аргументації з узагальненням, як це реалізовано у фреймворку AseSum. Цей підхід виходить за межі простого аналізу настрою ("батарея погана") до вилучення аргументів ("батарея погана, тому що вона деградує через 6 місяців").

AseSum використовує структуровану схему: твердження (аспект є позитивним/негативним), головна передумова (тип доказу) та другорядна передумова (спостереження користувача) [15].

Проектуючи відгуки на ці структури аргументації, система може генерувати резюме, які пояснюють причини консенсусу, надаючи споживачеві набагато багатшу відповідь на питання "Чому?" [16].

1.3 Метрики та бенчмарки

1.3.1 ROUGE та BERTScore

Можливо, найбільш значущою подією останніх років є "Криза оцінювання". Дослідницька спільнота значною мірою дійшла згоди, що традиційні n-грамні метрики є недостатніми для оцінювання абстрактивного узагальнення думок.

ROUGE вимірює перекриття n-грам (слів/фраз) між згенерованим резюме та еталоном, створеним людиною. В узагальненні думок модель може згенерувати валідне перефразування ("Батарея тримає весь день"), яке має нульове перекриття з еталоном ("Відмінне управління живленням"), що призводить до низької оцінки ROUGE попри високу якість.

BERTScore вимірює семантичну схожість за допомогою ембеддингів. Хоча вона краща за ROUGE, це метрика типу "чорна скринька", яка погано корелює зі специфічними помилками, такими як галюцинації або пропущені плюси/мінуси [22].

1.3.2 Парадигма "LLM-як-суддя"

Поточний протокол оцінювання SOTA передбачає використання високопродуктивної LLM (наприклад, GPT-4) як судді для вихідних даних моделі узагальнення. Ця парадигма дозволяє здійснювати нюансоване оцінювання, подібне до людського, у великих масштабах.

G-Eval (Generative Evaluation) використовує стратегію промптингу "Ланцюжок думок" (Chain-of-Thought, CoT). Судді-LLM надається вихідний

текст і згенероване резюме, і ставиться завдання оцінити його за шкалою 1–5 на основі конкретних вимірів [23].

1. Зв'язність: чи логічно структуроване резюме?
2. Узгодженість: чи відповідає воно джерелу?
3. Плавність: чи є граматики природною?
4. Релевантність: чи відображає воно найважливіші моменти?

Дослідження показують, що G-Eval досягає кореляції Спірмена 0.514 з людськими судженнями, значно перевершуючи традиційні метрики.

Оцінка QAG (Question Answer Generation) – це скоринг, який використовує високі можливості міркування LLM для надійної оцінки їхніх вихідних даних. Він використовує обмежені відповіді, зазвичай "так" або "ні", на закриті питання, які можуть бути згенеровані або задані заздалегідь, для обчислення фінальної метрики. Цей метод надійний, оскільки не використовує LLM для прямої генерації оцінок.

Наприклад, для обчислення оцінки правдивості (faithfulness), яка вимірює, чи є вихідні дані LLM галюцинацією, необхідно таке, перелічене нижче.

1. Використати LLM для вилучення всіх тверджень, зроблених у вихідних даних.
2. Для кожного твердження запитати у джерела істини, чи погоджується воно ("так") чи ні ("ні") зі зробленим твердженням.
3. Правдивість можна розрахувати як частку правдивих тверджень до загальної кількості тверджень, зроблених LLM [27].

Оскільки QAG не використовує LLM для прямої генерації балів оцінювання, але все ще використовує їхню здатність до міркування, це дозволяє отримати оцінки, які є і точними, і надійними.

1.3.3 Таксономія та виявлення галюцинацій

Галюцинація – генерація неправдивої інформації – є критичною проблемою для PRS.

Галюсії бувають двох видів: внутрішня та зовнішня. Внутрішня галюцинація - це коли резюме суперечить джерелу (наприклад, Джерело: "Батарея слабка", Резюме: "Чудова батарея"). Зовнішня галюцинація - коли резюме додає інформацію, відсутню в джерелі (наприклад, додавання "Водонепроникний", коли це не згадувалося).

Для тестування моделей проти цих типів помилок були розроблені бенчмарки, такі як FaithBench, та метрики, такі як Entity Hallucination Index (ЕНІ).

1.3.4 Калібрування полярності (PoSa)

Специфічною проблемою в узагальненні думок є дрейф тональності або упередженість полярності, коли резюме може бути більш позитивним або негативним, ніж фактичний розподіл відгуків. Калібрування полярності використовує навчання з підкріпленням (Reinforcement Learning, RL), щоб винагороджувати модель не лише за якість тексту, але й за відповідність розподілу тональності вихідного резюме вхідним відгукам. Це гарантує, що резюме точно відображає консенсус [29].

1.4 Порівняльний аналіз підходів та метрик

В наведених нижче таблицях описано ключові відмінності між обговореними архітектурами (табл. 1.1) та фреймворками оцінювання (табл. 1.2).

Таблиця 1.1 – Порівняльний аналіз архітектур узагальнення SOTA

<i>Архітектура</i>	<i>Філософія ядра</i>	<i>Інтеграція даних</i>	<i>Ключовий механізм</i>	<i>Найкращий сценарій використання</i>
MEDOS	Мультимодальне злиття	Відгуки + Описи + QA	Окремі енкодери та дистиляція мета-знань	Продукти "холодного старту"; Категорії зі складними специфікаціями (Електроніка).
FARSum	Усвідомленість зворотного зв'язку	Відгуки (фільтровані за метаданими)	Зважування за новизною та корисністю; Структурований вивід	Популярні продукти з великим обсягом даних;
XL-OPSUMM	Інкременталізм зі збереженням стану	Потокові відгуки	Словник аспектів для вирішення конфліктів та відстеження стану	Масивні платформи e-commerce з безперервними потоками відгуків.
AseSum	Аргументація	Відгуки + Схема аргументів	Вилучення структур Твердження-П ередумова	Сценарії, що вимагають глибокого пояснення ("Чому це погано?").
RAG + LLM	На основі пошуку	Витягнуті фрагменти (Chunks)	Гібридний пошук (Векторний/Ключові слова)	Гнучке узагальнення під запит; Ad-hoc запити з низькою затримкою.

Таблиця 1.2 – Порівняльний аналіз метрик оцінювання

<i>Метрика</i>	<i>Тип</i>	<i>Що вимірює</i>	<i>Кореляція з людським судженням</i>	<i>Плюси</i>	<i>Мінуси</i>
ROUGE	n-грама	Лексичне перекриття	Низька	Стандартний базовий рівень; Швидка; Дешева.	Штрафує валідні перефразування; Пропускає галюцинації.
BERTScore	Ембеддинг	Семантична схожість	Середня	Відображає зміст краще за ROUGE.	"Чорна скринька"; Не виявляє дрібних фактичних помилок.
G-Eval	На основі LLM	Зв'язність, Плавність, Релевантність	Висока	Імітує людську інтуїцію; Гнучкі критерії.	Дорога; Повільна; Потенційна упередженість LLM.
QAG	На основі LLM	Достовірність, Повнота	Дуже висока	Діагностична; виявляє помилки на рівні речень.	Висока обчислювальна вартість.
ЕНІ	На основі сутностей	Рівень галюцинацій	Висока	Виявляє помилки конкретних сутностей (бренди, специфікації).	Вузький фокус (тільки сутності).

1.5 Набори даних та бенчмарки

Валідність будь-якого оцінювання обмежена якістю тестових даних. В аспектно-орієнтованому узагальненні (ABSS) створення наборів даних розділяється на два підходи: високоякісна анотація людиною, яка є дорогою та рідкісною, та великомасштабне синтетичне конструювання, яке є

зашумленим, але доступним. Надійна стратегія оцінювання повинна використовувати обидва підходи для тестування різних можливостей моделі.

SPACE представляє поточний рівень SOTA у бенчмарках оцінювання для цієї області. На відміну від попередніх наборів даних, що покладалися на евристику, SPACE було створено за протоколом суворої анотації людиною, спеціально розробленим для тестування аспектно-орієнтованих можливостей [32].

1. Домен і масштаб: набір даних включає відгуки про готелі з TripAdvisor. Цей домен ідеально підходить для ABSS, оскільки відгуки про готелі природно структуровані навколо чітких, повторюваних аспектів, таких як Обслуговування, Чистота, Розташування, Їжа, Будівля та Номери.
2. Вхідні дані "Big Data": визначальною особливістю SPACE є масштаб вхідних даних. Кожна сутність у тестовому наборі пов'язана зі 100 відгуками. Це віддзеркалює реальний виклик, де узагальнювач повинен переварити величезну кількість тексту, що значно перевищує контекстні вікна багатьох стандартних моделей, змушуючи систему виконувати ефективний відбір контенту та кластеризацію [33].
3. Мультиреференсні золоті резюме: для 50 сутностей в оцінювальному наборі анотатори написали резюме для кожного з шести аспектів, плюс загальне резюме. Критично важливо, що кожне аспектне резюме має три різні людські референси. Цей мультиреференсний підхід є життєво важливим для оцінювання, оскільки він враховує варіативність того, як різні люди можуть узагальнювати одні й ті самі думки, тим самим стабілізуючи метрики, такі як ROUGE та BERTScore [34].
4. Стратегічна корисність: SPACE є основним бенчмарком для оцінювання Контрольованості. Оскільки він надає окремі базові

істини для "Чистоти" та "Обслуговування" для одного готелю, можна перевірити, чи модель, якій доручено узагальнити "Чистоту", випадково не включить інформацію про "Обслуговування" у свій вивід – поширений режим відмови, відомий як сплутування аспектів [35].

Хоча SPACE забезпечує якість, йому бракує масштабу, необхідного для навчання глибоких моделей з учителем. Щоб подолати цей розрив, дослідники розробили *OPOSUM* та його розширеного наступника *OPOSUM+* [36].

Ці набори даних використовують техніку "слабкого нагляду" або "синтетичної" генерації. Автори беруть великий корпус відгуків Amazon і розглядають окремий відгук як цільове резюме. Потім вони шукають у корпусі інші відгуки про той самий продукт, які семантично схожі на цей цільовий відгук. Ці знайдені відгуки стають "вхідними документами". Це створює тренувальну пару, вхідний набір та цільовий відгук, без ручного написання [37].

OPOSUM+ є критичним для тестування узагальнення домену. Він охоплює шість різних категорій продуктів: сумки для ноутбуків, bluetooth-гарнітури, черевики, клавіатури, телевізори та пилососи. Це дозволяє дослідникам оцінити, чи може модель, навчена на електроніці, адаптувати свої механізми ідентифікації аспектів до одягу [34].

Покладання на існуючі відгуки користувачів як на еталонні дані вносить специфічну упередженість: розповідь від першої особи. Відгуки користувачів часто містять фрази на кшталт "Я купив це для свого сина". Справжнє резюме повинно бути від третьої особи та узагальнюючим, наприклад, "Користувачі повідомляли про повернення товару...". Моделі, інтенсивно навчені на OPOSUM+, часто не можуть прийняти правильний "голос" узагальнення, імітуючи окремого користувача замість агрегатора [38].

AmaSum вирішує проблему "голосу" OPOSUM+, використовуючи професійні резюме [39].

1. Джерело: він поєднує відгуки про продукти Amazon з резюме, написаними професійними критиками з великих оглядових вебсайтів.
2. Структура: цільові резюме є високоструктурованими, розділеними на секції Вердикт, Плюси та Мінуси. Така структура робить AmaSum унікально придатним для навчання моделей генерувати структуровані вихідні дані.
3. Масштаб: маючи понад 31 000 продуктів, це один із найбільших доступних наборів даних.
4. Упередженість: критичним застереженням щодо AmaSum є те, що професійний критик часто пише своє резюме на основі власного досвіду використання продукту, а не виключно на основі відгуків користувачів. Це створює пастку галюцинацій під час оцінювання: модель може бути оштрафована за негенерування факту, присутнього в резюме критика, який ніколи не згадувався у відгуках користувачів [39].

Більшість систем передбачають фіксовану онтологію аспектів (наприклад, 6 аспектів у SPACE). Однак у реальному світі постійно з'являються нові аспекти. *OpenAsp* тестує узагальнення на основі відкритих аспектів (OABS) [40]. Цей набір даних оцінює здатність моделі ідентифікувати та узагальнювати довільні аспекти, визначені під час інференсу, виходячи за межі обмежень фіксованих класифікаторів аспектів.

У той час як SPACE представляє точність, а OPOSUM – синтетичний обсяг, набір даних *Wayfair Customer Reviews* представляє реальні відгуки з одного з найбільших комерційних майданчиків. Випущений разом зі статтею End-to-End Aspect-Guided Review Summarization at Scale (EMNLP 2025

Industry Track), він є сериодною між академічними бенчмарками та реальним впровадженням в e-commerce [41].

1. Домен і масштаб: набір даних складається з реальних відгуків клієнтів з Wayfair.com, зосереджуючись на домені товарів для дому та побуту.
2. Масивний масштаб вхідних даних: це наразі один із найбільших публічних наборів даних для цього завдання, що містить приблизно 11,8 мільйона відгуків та 91 000 резюме на рівні продукту. На відміну від SPACE, Wayfair пропонує "макро" погляд, дозволяючи моделям навчатися на величезному шумі та варіативності мільйонів унікальних продуктів.
3. Гранулярне аспектне керівництво: ключовою інновацією є включення дрібнозернистих міток аспектів на рівні відгуку. Набір даних надає JSON-форматовані анотації для аспектів (наприклад, {"Міцність": "POSITIVE", "Збірка": "NEGATIVE"}), поєднані з сирим текстом. Ця подвійна структура дозволяє Багатозадачне навчання: модель можна навчити одночасно виконувати аспектно-орієнтований аналіз тональності (ABSA), щоб ідентифікувати, що обговорюється, та Узагальнення для генерації плавного огляду цих тем.

В таблиці 1.3 наведено порівняння основних наборів даних.

Таблиця 1.3 – Стратегічне порівняння основних наборів даних для оцінювання

<i>Набір даних</i>	<i>Джерело нагляду</i>	<i>Структура аспектів</i>	<i>Масштаб (Тестові сутності)</i>	<i>Розмір входу (Відгуків / Сутність)</i>	<i>Основна корисність оцінювання</i>
SPACE	Експерт (Золотий стандарт)	Фіксована (6 типів)	50	100	Точність / контроль: тестування суворого дотримання аспектів та точності тональності.
OPOSUM +	Синтетичний (Відгуки користувачів)	Фіксована (6 доменів)	60 на домен	~10 (Синтетичний пошук)	Узагальнення: тестування ефективності в різних категоріях продуктів.
AmaSum	Професійні критики	Структурована (Плюси / Мінуси)	50 на домен	~320	Структура та масштаб: обробка масивних вхідних даних та генерація форматованих списків.
OpenAsp	Людська екстракція	Відкрита / Динамічна	1,310 (Всього)	Різна (Multi-doc)	Гнучкість: тестування ідентифікації динамічних, нестандартних аспектів.
Wayfair	Слабкий (Згенеровано LLM)	Гранулярна / Динамічна	~91,000	~129 (Серед) / 200 (Макс)	Тестування стійкості до шуму та масштабу в реальних даних e-commerce.

Висновки до розділу 1

За останні роки, траєкторія розвитку узагальнення відгуків про товари підкреслює чітку міграцію від узагальнення тексту до інтелектуального аналізу думок. Найуспішнішими системами є ті, що враховують унікальні властивості відгуків: їхню суб'єктивність, зашумленість, тимчасовість та структурованість.

РОЗДІЛ 2 ЗАПРОПОНОВАНИЙ ПІДХІД

Основною метою цього дослідження є трансформація неструктурованого, зашумленого та об'ємного зворотного зв'язку користувачів у кількісне аспектно-орієнтоване узагальнення (Quantified Aspect-Based Summary, QABS) – структуровану, неупереджену та придатну до дії агрегацію думок користувачів.

Сучасні методи узагальнення часто покладаються на промптинг великих мовних моделей (LLM) як на "чорну скриньку", що може призводити до галюцинацій, відсутності покриття специфічних доменів та неперевірюваних тверджень.

Для подолання цих обмежень ми пропонуємо модульний, багатоетапний пайплайн, розроблений для перетворення сирих відгуків у структуровані атомарні твердження перед їх синтезом.

Цей розділ визначає цільовий вивід (QABS), деталізує запропоновану лінійну архітектуру та встановлює метрики оцінювання для забезпечення критеріїв якості резюмування відгуків.

2.1 Визначення цілі: кількісне аспектно-орієнтоване узагальнення (QABS)

Перед визначенням архітектури необхідно встановити критерії якісного резюме. Ми пропонуємо цільовий стандарт резюмування відгуків - кількісне аспектно-орієнтоване узагальнення (Quantified Aspect-Based Summary, QABS). На відміну від загального скорочення тексту, QABS визначається як структурована, неупереджена та дієва агрегація відгуків користувачів,

розроблена для забезпечення впевнених рішень про покупку. Приклади QABS наведені в додатку А.

2.2.1 Критерії якості

Ми визначили сім основних критеріїв, яким повинно відповідати резюмування відгуків.

1. **Ясність та доменна експертиза:** резюме має демонструвати спеціалізовані знання про категорію продукту, наприклад, розрізнення між дзеркальними та бездзеркальними камерами, та використовувати правильну технічну лексику.
2. **Практичність:** воно повинно артикулювати основні компроміси (Плюси/Мінуси) та надавати чіткі рекомендації щодо покупки, адаптовані до різних профілів користувачів, наприклад, "Кому це варто купувати?".
3. **Покриття та категоризація:** система повинна агрегувати сентимент за всіма релевантними категоріями, включаючи похвалу продукту, дефекти, критику продуктивності та цінність.
4. **Верифікованість:** ключові твердження мають бути підкріплені прямими, репрезентативними цитатами з атрибуцією джерела для запобігання галюцинацій.
5. **Квантифікація та зважування:** тональність повинна бути представлена кількісно, наприклад, "75% користувачів...", забезпечуючи статистичне підґрунтя для резюме.
6. **Структурованість:** вивід повинен використовувати чіткі заголовки, марковані списки та таблиці для максимальної читабельності.

7. Якість написання.

- a. Плавність: чи є граматики та правопис правильними?
- b. Зв'язність: чи є текст логічно послідовним?
- c. Релевантність: чи відповідає він на запит (промпт)?
- d. Фактологічна узгодженість: чи є інформація правдивою на основі джерела?

2.2.2 Структура даних

Щоб гарантувати дотримання цих критеріїв, ми розглядаємо резюме як об'єкт структурованих даних, а не як простий текстовий рядок. Запропонована цільова структура сегментує відгук на вступ, основну частину, що містить мінуси, плюси та корисну інформацію, та висновок.

JSON

```
{ProductReviewSummary: {
  {Introduction}
  {Body:
    [{Cons: [aspectsReview]},
    {Pros: [aspectsReview]},
    {Info: [aspectsReview]}}]
  {Conclusion: [recommendations]}
}}
```

де:

JSON

```
aspectsReview = {
  aspectName: {
    aspectReviewSummary,
    aspectReviewInsight: {aspectReviewQuote, reviewId},
```

```

        aspectSentimentStats
    }
}

```

Ця ієрархічна структура змушує генеративну модель наводити до кожного узагальнення аспекту цитату (`aspectReviewQuote`) та ідентифікатор відгуку (`reviewId`) для його підтвердження, а також наводити статистичне підґрунття (`aspectSentimentStats`), що безпосередньо відповідає критеріям верифікованості та квантифікації.

2.3 Порівняння з існуючими реалізаціями резюмування

Автоматично згенеровані резюме відгуків про товари є швидко зростаючою потребою, проте багато технологічних гігантів пропонують обмежені рішення.

Review Highlights групує відгуки користувачів навколо згаданого аспекту, такого як якість їжі або обслуговування. Наприклад, Yelp та Youtube впровадили такі підходи на своїх платформах. Порівняно з нашим запропонованим стандартом узагальнення, цьому підходу бракує читабельності та зв'язності, які може забезпечити повноцінне резюме.

Абстрактивне узагальнення із заздалегідь визначеними аспектами. Надаючи абстрактивне резюме аспектів відгуків про продукт, цей метод забезпечує користувачам аналіз відгуків та їхню тональність щодо аспектів продукту.

TripAdvisor Review Summaries пропонує заздалегідь визначений набір з 5-6 аспектів щодо апартаментів.

Review Highlights



"This was our first meal in Madrid, we got to our hotel from the airport and wanted some **tapas** and WOW." in 10 reviews



"Our main courses were **Pulpo** a la Plancha (grilled octopus) and Gambones Al Ajillo (garlic shrimp)." in 4 reviews



"We enjoyed the **ropa vieja**, garlic shrimp and the grilled vegetables with goat cheese." in 5 reviews

Рисунок 2.1 – Yelp review highlights [42]

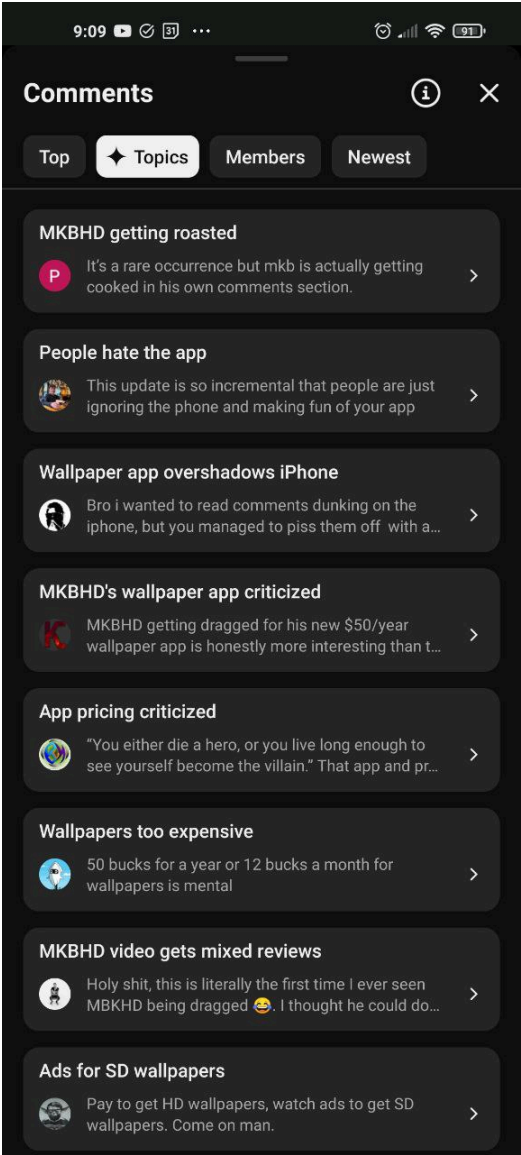


Рисунок 2.2 –Youtube comment topics [43]

Reviews summary

This summary was created by AI, based on recent reviews.

Location

Central

Rooms

Comfortable

Cleanliness

Spotless

Service

Attentive

Atmosphere

Welcoming

Value

Excellent

Noise level

Quiet

Amenities

Well-appointed

La Siesta Classic Ma May Hotel is popular among travelers for its excellent value and spotless cleanliness. The rooms, often praised for their comfort and spaciousness, provide a quiet retreat despite the lively Old Quarter location. Guests appreciate the central and convenient setting, which allows easy exploration of Hanoi.

The warm and attentive service makes visitors feel well cared for. Amenities, including a rooftop bar and relaxing spa, earn rave reviews, while the stylish atmosphere adds to the overall experience.

Jump to all reviews

Рисунок 2.3 – TripAdvisor review summaries [44]

Кожен аспект, наприклад, розташування, має власне резюме.

Location

Atmosphere

Rooms

Value

Cleanliness

Noise level

Service

Amenities

Location

Central

For many travelers, the hotel's location earns rave reviews for its convenience and centrality in Hanoi's Old Quarter, making it easy to explore attractions, restaurants, and markets by foot. Some note it's lively but still peaceful inside.

Ragesh Nair

Jul 2025 • 1 contribution

...Since the hotel is located in Old Quarter you may find the basic room size to be smaller (still value for money considering the location) and I guess it can be the same in most of the hotels in Old...

See full review

FarAway13949138347

Aug 2025 • 1 contribution

...The location is unbeatable—right in the heart of the Old Quarter, making it easy to explore Hanoi's vibrant streets, markets, and eateries...

See full review

Hyun Jun Park

Jul 2025 • 1 contribution

...Breakfast is very delicious, so you can not go out in the morning, and the hotel location is good, so you can walk around the restaurants, cafes, night markets, and the sea is fine, so it was goo...

See full review

Рисунок 2.4 – TripAdvisor (aspects) review summaries [44]

Кожен аспект перераховує відгуки, де він згадується. При натисканні на відгук в аспекті відкривається повний відгук з виділеним фрагментом, де згадується аспект.

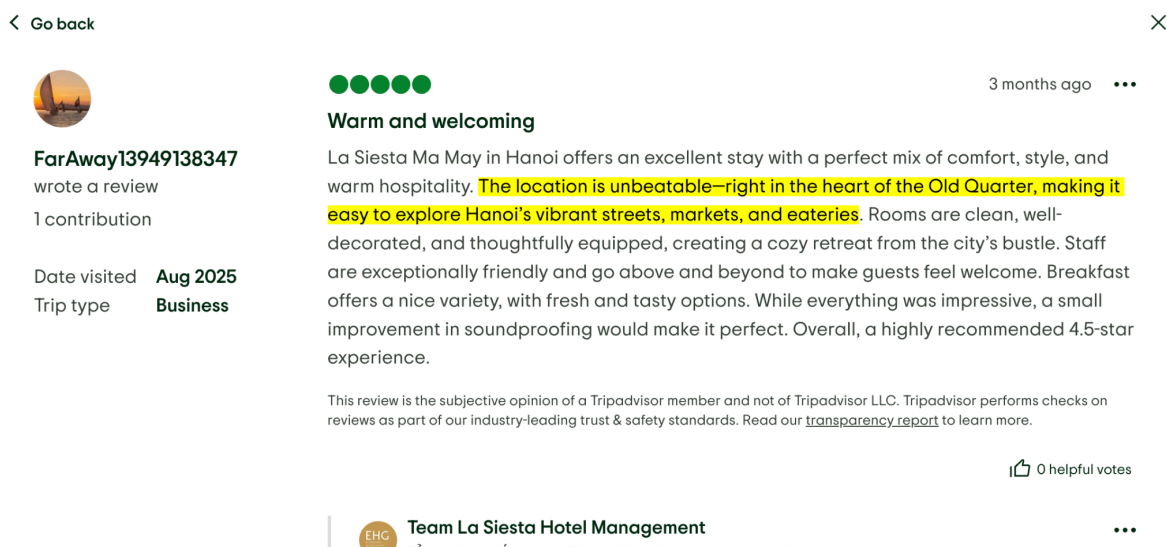


Рисунок 2.5 – TripAdvisor review [44]

Порівняно з нашим підходом, цим резюме бракує гнучкості у виборі аспектів, що призводить до виключення багатьох важливих тем, піднятих у відгуках, роблячи саме резюме менш інформативним та більш упередженим. Крім того, наш підхід надає статистику тональності аспектів для кращої верифікованості.

Абстрактивне узагальнення на основі запиту з динамічними аспектами. Reddit Answers надає відповіді на запитання користувачів резюмуючи коментарі, що належать до відповідної теми.

Is Gollum from the Lord of the Rings an evil or a good character?



Sources: r/changemyview, r/lotr, r/lordoftherings +2 more >

Gollum, also known as Sméagol, is one of the most complex and debated characters in "The Lord of the Rings." Here's a succinct guide to answer whether he is considered evil or good, based on the opinions of various Redditors:

Early Character and Influence of the Ring

- **Initial Actions:** Sméagol's descent into darkness began before he even touched the Ring. He murdered his friend Deagol to obtain it, indicating a pre-existing inclination towards greed and violence. "[Smeagol murdered his friend to get the Ring.](#)" >
- **Immediate Corruption:** The Ring's influence was immediate and profound, turning him from a simple Hobbit into a twisted creature. "[He killed Deagol moments after seeing the ring, in the movies anyway.](#)" >

Dual Personalities

- **Sméagol vs. Gollum:** The character is often seen as having two distinct personalities—Sméagol, representing his former self, and Gollum, the corrupted version. "[Smeagol is broken. Gollum is pure evil.](#)" >

Рисунок 2.6 – Reddit Answers [45]

На відміну від інших підходів, цей метод пропонує дуже зручний спосіб для користувача аналізувати відгуки про продукт за допомогою запиту. Більше того, аспекти продукту, що з'являються в резюме, не є заздалегідь визначеними, тому надають повну картину продукту, відповідно до запиту. Для верифікованості Reddit Answers надає цитату з відгуку до кожного резюме аспекту з посиланням на оригінальний відгук.

У порівнянні з нашим підходом, цей метод досягає кращої релевантності завдяки генерації резюме, яке відповідає на запит користувача. Однак, з точки зору верифікованості, йому бракує статистичного підґрунтя для підтвердження сентименту аспекту.

2.4 Архітектура системи

Запропонована архітектура є лінійним, багатоетапним пайплайном. На відміну від end-to-end LLM підходів, ця система трансформує неструктурований текст у структуровані дані (атомарні твердження або кластери) перед тим, як конвертувати його назад у синтезований текст. Така модульність дозволяє здійснювати спеціалізовану оптимізацію на кожному етапі.



Рисунок 2.7 – Apple App Store Review Highlights [46]

Етап 1: фільтрація вхідних даних. Перший етап забезпечує якість даних шляхом видалення шуму. Автоматизовані фільтри застосовуються до потоку відгуків для усунення спаму, ненормативної лексики та шахрайських відгуків. Цей крок є критичним для запобігання посиленню шкідливого або нерелевантного контенту у фінальному резюме.

Етап 2: вилучення інсайтів (Insight Extraction). Цей етап декомпонує складні, розмовні відгуки на "інсайти" – атомарні твердження, обмежені однією темою та тональністю.

Фаза 1: вирішення кореферентності. Сирі відгуки часто використовують займенники, які приховують зміст, якщо розглядати речення ізольовано. Наприклад, у реченні "Я купив S24. Він швидкий" стандартний екстрактор може не пов'язати "Він" з "S24". Ми використовуємо модель вирішення кореферентності (Coreference Resolution), наприклад, SpanBERT, для

вирішення цих залежностей перед вилученням, трансформуючи "Він швидкий" у "S24 швидкий".

Фаза 2: вилучення кортежів за допомогою LoRA. Ми використовуємо велику мовну модель на 7 млрд параметрів, наприклад, Llama-3, донавчану за допомогою низькорангової адаптації (LoRA) на наборах даних аспектно-орієнтованого аналізу тональності (ABSA). Модель отримує інструкцію вилучати JSON-об'єкт (кортеж) для кожної окремої думки.

Вхід: "The battery drains fast, but the camera is amazing for portraits."

Вихідні кортежі:

JSON

```
{Aspect: "Battery", Sentiment: "Negative", Intensity: 0.9,
Descriptor: "drains fast", Quote: "battery drains fast"}
{Aspect: "Camera", Sub-Aspect: "Portraits", Sentiment:
"Positive", Intensity: 1.0, Descriptor: "amazing", Quote:
"camera is amazing for portraits"}
```

Етап 3: динамічне моделювання тем. Цей етап агрегує атомарні твердження для динамічного виявлення тем, не покладаючись на заздалегідь встановлену таксономію.

Фаза 1: генерація назв тем. Друга LoRA-налаштована модель аналізує вилучені інсайти та генерує динамічні й описові мітки для нових кластерів.

Фаза 2: дедуплікація. Щоб забезпечити лаконічність списку тем, ми використовуємо два методи.

1. Семантичні ембеддинги: інсайти проєктуються у векторний простір. Щільні кластери ідентифікуються як єдині теми. Наприклад, "не можу увійти" та "помилка аутентифікації" об'єднуються на основі семантичної близькості, незважаючи на різне формулювання.
2. Пошук за шаблоном: використання regex патернів для пошуку незначних варіацій в написанні або фразуванні.

Етап 4: відбір тем. Не всі знайдені теми варто вносити до фінального резюме. Ми застосовуємо алгоритм відбору на основі п'яти конкретних метрик:

1. Релевантність: щоб розрізнити теми, що належать до "Теми про товар" від "Теми поза товаром", ми класифікуємо теми на основі інформації про продукт шляхом RAG підходу. Ми надаємо пріоритет темам, що стосуються функцій, продуктивності та дизайну товару, тоді як теми поза товаром, наприклад, думки про якість їжі у відгуку про застосунок для доставки їжі, депреорітизується. Клас "Теми про товар" має два підкласи: "Основний аспект" та "Додатковий аспект". До основних аспектів входять фундаментальні теми про товар, які мають бути у резюме навіть якщо кількість відгуків про них не чисельна. До додаткових аспектів входять усі не основні аспекти. Такий поділ дозволяє охопити всі базові аспекти, надаючи їм найвищу релевантність.
2. Популярність: щільність кластера = (Згадування / Всього відгуків) * 100.
3. Актуальність: надання більшої ваги недавнім інсайтам для відображення поточної версії додатку.

$$W_t = W_0 * e^{-\lambda t},$$

де t – вік відгуку.

4. Показник сентименту: середня оцінка тональності.
5. Показник суперечливості: висока дисперсія тональності вказує на поляризовані погляди на аспект.

Етап 5: відбір інсайтів. З відібраних тем система вибирає репрезентативні атомарні твердження, які слугуватимуть текстовою основою для резюме. Ми використовуємо репрезентативну вибірку (Representative Sampling), обчислюючи топ-N інсайтів, найближчих до центроїди кластера аспекта.

Етап 6: генерація резюме. Фінальний етап синтезує відібрані інсайти в цілісну розповідь. Ми уникаємо загального промптингу, використовуючи Blueprint Prompt. Система створює чіткий набір інструкцій для LLM: "Напиши розділ під назвою '## Плюси'. Використай наступні 3 структуровані інсайти. Для Інсайту 1 (Батарея) використовуй статистику '75%'. Використай ID цитати...". Це змушує LLM діяти як соціолог, обмежений фактами, а не як креативний генератор.

2.5 Порівняння з існуючими архітектурами

Запропонована архітектура має кілька переваг і недоліків порівняно з усталеними підходами.

Переваги:

- масштабування даних: завдяки ABSA (етапи 2-3), що використовує логіку map-reduce, система здатна обробляти великі дані (Big Data), на відміну від LLM, які обмежені розміром контекстного вікна;
- мікросервісна архітектура: дозволяє ефективне масштабування системи;
- модульність: пайплайн сприяє модельно-агностичному підходу та гнучкий до внесення модифікацій;
- динамічне моделювання тем: досягає кращої релевантності та покриття, ніж заздалегідь визначений набір аспектів.
- контекстуалізація: покращена обізнаність завдяки використанню RAG при відборі аспектів.

Недоліки:

- каскадна композиція помилок, спричинена запропонованою лінійною архітектурою, створює ланцюг залежностей, де помилки накопичуються від рівня до рівня;
- інженерна складність: підхід вимагає створення та підтримки багатьох компонентів (фільтри, кластеризатори, екстрактори), а не простого управління одним промптом.

2.6 Методологія оцінювання

Для валідації запропонованого підходу ми вводимо набір нових метрик оцінювання, розроблених для прямого відображення критеріїв QABS.

1. Ясність та доменна експертиза – TF-IDF лексики, специфічної для предметної області.
2. Практичність – пошук за шаблоном на наявність в резюме рекомендацій, застережень та порівнянь.
3. Покриття та категоризація – відсоток значущих аспектів із джерела, що покриті в резюме.
4. Верифікованість – відсоток використаних цитат, що за відстанню Левенштейна мають достатню схожість із першоджерелом.
5. Квантифікація та зважування – середня абсолютна похибка (MAE) між заявленими розподілами тональності в резюме та дійсним розподілом у джерелі.
6. Структурованість – пошук за шаблоном на наявність заголовків, списків та виділень жирним шрифтом.
7. Якість написання.

- a. Плавність та зв'язність – оцінювання за методом LLM-як-суддя.
- b. Релевантність – відсоток охоплених підтем в резюмуванні кожного аспекту
- c. Фактологічна узгодженість – відповідність тональності в резюмуванні аспектів до статистики їхньої тональності.

Однак у цій роботі ми зосередилися лише на чотирьох з них: покриття та категоризація, верифікованість, квантифікація та зважування, та якість написання.

Покриття та категоризація. Щоб гарантувати, що резюме є всебічним, ми пропонуємо метрику *повноти семантичного покриття аспектів* (Semantic Aspect Coverage Recall, SACR). Вона вимірює, чи покриває резюме повний набір аспектів – як статистично значущих тем (A_{sig}), так і обов'язкових для домену аспектів (A_{req}) – використовуючи семантичну схожість.

$$SACR = \frac{|A_{sum} \cap_{semantic} A_{src}|}{A_{src}},$$

де

- A_{src} (цільовий набір аспектів): об'єднання статистично популярних тем та обов'язкових доменних тем.

$$A_{src} = A_{sig} \cup A_{req},$$

де

- A_{sig} (значущі аспекти): аспекти, вилучені з вихідних відгуків, що з'являються у >K% речень.

- A_{req} (обов'язкові аспекти): підмножина заздалегідь визначених аспектів (A_{def}), які фактично існують у вихідному тексті (A_{all}).

$$A_{req} = A_{def} \cap_{semantic} A_{all}.$$

Це гарантує, що модель не штрафується за пропуск стандартного аспекту, наприклад, "Водонепроникність", якщо користувачі про нього не згадували.

- $\cap_{semantic}$ (семантичний перетин): на відміну від стандартного перетину множин, аспект $a \in A_{src}$ вважається покритим, якщо він має достатню семантичну схожість з яким-небудь аспектом у резюме A_{sum} .

●

$$I_{covered}(a) = \{1 \text{ if } \max_{s \in A_{sum}} (\cos(E(a), E(s))) > \tau \text{ else } 0\},$$

де

- $E(\cdot)$ - функція ембедингу речень, як от SBERT
- τ - заданий поріг подібності.

Верифікованість. Для запобігання галюцинаціям (сфабрикованим цитатам або фактам) у резюме, ми пропонуємо метрику *рівень галюцинацій у цитатах* (Quote Hallucination Rate, QHR), яка обчислює відсоток цитат у резюме, що відсутні у джерелі.

$$QHR = \frac{N_{fabricated}}{N_{total}}.$$

Система перевіряє, чи існує зазначений reviewId. Якщо так, вона виконує нечітке зіставлення рядків (відстань Левенштейна) між цитатою в

резюме та вихідним реченням. Якщо подібність нижча за поріг τ (наприклад, 90%), цитата вважається сфабрикованою.

Квантифікація та зважування. Щоб гарантувати, що резюме не перебільшує і не применшує тональність користувачів, запобігаючи статистичним галюцинаціям, ми пропонуємо метрику *узгодженості статистичних тверджень* (Statistical Claim Consistency, SCC). Вона вимірює середню абсолютну похибку (MAE) між відсотками, заявленими в резюме, та розподілом тональності істинного значення (ground-truth), розрахованим з вихідного тексту.

Алгоритм.

1. Розрахунок істинного значення: для кожного аспекту в резюме, класифікувати тональності на всіх реченнях вихідних відгуків, що містять цільовий аспект. Розрахувати фактичне співвідношення позитивних/негативних відгуків.
2. Верифікація: порівняти вилучений відсоток з резюме з розрахованим істинним значенням.

$$SCC = 1 - \frac{1}{N} \sum_{i=1}^N \frac{|Claimed_i - Actual_i|}{100}.$$

Якість написання. Щоб переконатися, що резюме фактично узгоджене та релевантне, для кожного аспекту, наведеного в резюме, мають бути перевірені його тональність та підтеми.

Для оцінки того, наскільки добре тональність резюме аспекту узгоджується з розподілом тональності аспекту, ми використовуємо метрику *узгодженість тональності резюме аспекту* (Aspect Summary Sentiment Alignment, ASSA). Вона розраховує тональність резюме аспекту та порівнює її з розрахованим розподілом тональності аспекту.

Алгоритм.

1. Розрахунок істинного значення: ідентично до першого кроку для розрахунку SCC , для кожного аспекту в резюме, класифікувати тональності на всіх реченнях вихідних відгуків, що містять цільовий аспект. Розрахувати фактичне співвідношення позитивних/негативних відгуків.
2. Визначення настрою резюме аспекту: класифікувати настрій кожного резюме аспекту в резюме.
3. Верифікація: порівняти настрої в резюме аспектів з розрахованим істинним значенням.

$$ASSA = 1 - \frac{1}{N} \sum_{i=1}^N \frac{|Summary_i - Actual_i|}{100},$$

де N – кількість наведених аспектів в резюме.

Для оцінки того, наскільки добре резюме аспекту покриває значущі підтеми, ми вводимо метрику *звážена повнота за ключовими словами тональності* (Weighted Sentiment-Keyword Recall, WSKR).

$$WSKR = \frac{\sum_{w \in (Summary \cap TopK)} Importance(w)}{\sum_{w \in TopK} Importance(w)},$$

де

- $TopK$ - це множина розміром з K найчастіше вживаних підтем в аспекті.

$$TopK = \arg_{S \subseteq W} \max_{|S|=k} \sum_{w \in S} importance(w).$$

- Importance(w) - функція важливості, яка визначає наскільки вагомою є підтема для сентимента аспекта.

$$importance(w) = \log(Frequency_w + 1) * |AvgSentiment_w - 0.5|$$

Для оцінки плавності та зв'язності ми пропонуємо дві альтернативи: оцінювати ефективність моделі на загальних завданнях генерації тексту, використовуючи класичні контрольовані набори даних і метрики ROUGE-L або COMET, або оцінювати ефективність моделі на основному завданні, використовуючи LLM-як-суддю. У цій роботі ми пропонуємо використати перший підхід, оскільки він є менш обчислювально витратним.

2.6 Порівняння з існуючими метриками

Головною перевагою запропонованих метрик є те, що для їхнього використання не потрібно мати еталонне резюме. Стандартні методи, такі як ROUGE та BERTScore, порівнюють згенеровані резюме з еталонними, які є дуже рідкісними через складність та вартість їх написання людиною для товарів, що мають сотні, а то й тисячі, відгуків.

Наші метрики порівнюють згенероване резюме з аспектно орієнтованим семантичним аналізом, зібраним з відгуків. Для їхньої роботи достатньо мати маркований датасет відгуків, як для ABSA задачі, що є значно простішим в розмітці, адже людині необхідно лише визначити аспекти та їхню оцінку у відгуку, а не аналізувати увесь датасет з подальшим написанням резюме.

При відсутності розмічених відгуків, можна використовувати існуючі натреновані ABSA моделі. Це дозволяє перейти від керованого навчання до

напівкерованого, що є дуже важливою перевагою в умовах нестачі розмічених даних.

Іншою сильною стороною запропонованих метрик є їхня низька обчислювальна складність у порівнянні з LLM-як-суддя підходами, як G-Eval та QAG.

Висновки до розділу 2

У цьому розділі було запропоновано стандарт резюмування, системну архітектуру та метрики оцінювання.

Новий стандарт для резюмування відгуків, QABS, було визначено за 7 критеріями й представлено шаблоном у JSON форматі. Для порівняння, були взяті існуючі резюме з одних з найбільших ІТ компаній в цьому сегменті.

В ході роботи, було також визначено модульну системну архітектуру для резюмування відгуків, детально описано її етапи та порівнянно з існуючими підходами.

Насамкінець, було визначено спеціальні метрики з метою якнайкращого дотримання QABS стандарту. Метрики були співставленні з сучасними метриками оцінювання резюме - ROUGE, BERTScore, G-Eval та QAG.

Таким чином, у цьому розділі було запропоноване й обґрунтоване рішення поставленої в цій роботі проблеми.

РОЗДІЛ 3 РОЗРОБКА СТАРТАП-ПРОЕКТУ “ReviewMind AI”

3.1 Опис ідеї проекту та технологічний аудит

В умовах стрімкої цифровізації глобальної економіки та експоненційного зростання обсягів електронної комерції, взаємодія між брендом та споживачем трансформується у безперервний потік даних. Основою запропонованого стартап-проекту під робочою назвою "ReviewMind AI" є створення інноваційної SaaS-платформи для автоматизованого, глибокого семантичного аналізу відгуків клієнтів. Проект спрямований на вирішення критичної проблеми сучасної індустрії рітейлу та виробництва товарів масового вжитку – нездатності ефективно обробляти та інтерпретувати величезні масиви неструктурованих текстових даних (Big Data), що надходять з десятків розрізнених каналів комунікації (маркетплейси, соціальні мережі, служби підтримки).

Зміст ідеї полягає у розробці аналітичної екосистеми, яка фундаментально змінює підхід до роботи з фідбеком: від реактивного моніторингу до проактивного управління якістю продукту та клієнтським досвідом. На відміну від існуючих рішень, які вимагають тривалого ручного налаштування таксономій або обмежуються простим підрахунком ключових слів, наша система використовує архітектуру великих мовних моделей (LLM) нового покоління з розширеним контекстним вікном. Це дозволяє реалізувати концепцію "безшовного розуміння" (seamless understanding), коли штучний інтелект самостійно виявляє аспекти продукту (aspect discovery) без попереднього навчання з учителем.

Система базується на трьох ключових технологічних перевагах, які формують її унікальну ринкову пропозицію.

1. **Необмежене контекстне вікно для обробки Big Data:**
використання ABSA моделей для кластеризації і фільтрації

відгуків, за подібною логікою до map-reduce, здатних обробляти від 128 тисяч до 1 мільйона токенів в одному запиті (наприклад, Gemini 1.5 Pro або модифіковані версії GPT-4o), у поєднанні з архітектурою RAG (Retrieval-Augmented Generation), дозволяє системі аналізувати не окремі відгуки ізольовано, а цілісні масиви даних за тривалі періоди часу [47]. Це забезпечує виявлення складних, часозалежних патернів (наприклад, деградація якості товару після зміни постачальника компонентів півроку тому), які неможливо помітити при традиційному підході.

2. **Автоматична генерація аспектів (Unsupervised Aspect Discovery):** система використовує індуктивний метод аналізу. Замість того, щоб змушувати користувача вручну створювати категорії (наприклад, "Доставка", "Ціна", "Упаковка"), алгоритм самостійно кластеризує семантичні вектори відгуків і формує динамічну онтологію проблем. Це дозволяє виявляти нестандартні, нішеві або нові проблеми (наприклад, "специфічний запах пластику при нагріванні" або "незручне розташування кнопки живлення для лівшів"), які не були передбачені стандартними класифікаторами [49]. Це критично важливо для швидкого реагування на дефекти нових продуктів.
3. **Економічна ефективність та помірна ціна:** завдяки використанню каскадної архітектури обробки запитів, де рутинні задачі виконуються оптимізованими "малими" моделями (SLM або GPT-4o-mini), а дорогі ресурсомісткі моделі залучаються лише для фінального синтезу складних інсайтів, вдається досягти значного зниження собівартості аналізу [51]. Це дозволяє пропонувати продукт за ціною, доступною для середнього

бізнесу (Mid-market), на відміну від дорогих Enterprise-рішень конкурентів.

Унікальна цінність проекту визначається можливістю отримати комплексний зріз думок споживачів за лічені хвилини без необхідності залучення дата-аналітиків, що знижує бар'єр входу для компаній, які раніше не могли дозволити собі глибоку аналітику. Система демонструє високу адаптивність до різних доменних областей – від електроніки та косметики до SaaS-продуктів та мобільних додатків.

Найбільш перспективним є ринок прямих продажів (Direct-to-Consumer, DTC), де репутація бренду та швидкість реакції на відгуки є критичними факторами виживання. Очікувані клієнти включають бренди, що масштабуються на платформах Shopify, Amazon, та Magento, а також продуктові IT-компанії. Ринок Customer Analytics, на якому оперує стартап, оцінюється у понад 17 мільярдів доларів США у 2024 році з прогнозованим зростанням до 48.6 мільярдів до 2030 року [52].

Аналіз конкурентного середовища демонструє, що існуючі рішення можна поділити на три категорії: платформи збору відгуків з базовою аналітикою (Yotpo), спеціалізовані інструменти текстового аналізу (MonkeyLearn, Keatext) та комплексні Enterprise-платформи управління клієнтським досвідом (Chattermill, Thematic, Qualtrics). Наш проект позиціонується як "золота середина", що поєднує глибину аналізу Enterprise-рішень з доступністю та простотою SaaS-інструментів для масового ринку.

В табл. 3.1 продемонстровано детальне порівняння техніко-економічних характеристик пропонованого стартапу з провідними гравцями ринку.

Таблиця 3.1 – Техніко-економічні характеристики ідеї проекту

Техніко-економічна характеристика	<i>Chattermill</i>	<i>Thematic</i>	<i>MonkeyLearn</i>	<i>Yotpo (Analytics)</i>	<i>Наш проект (ReviewMind AI)</i>
Вартість (\$/місяць)	Висока (Custom Enterprise Pricing), часто від \$1000+	Висока (від \$2000/міс за базовий пакет) [61]	Середня (\$299 - \$1200+), оплата за обсяг	Входить у дорогі пакети (Premium/Enterprise), від \$500+	Помірна (\$199 - \$799), прозора SaaS модель
Метод виявлення тем	Unsupervised + Supervised (вимагає навчання та налаштування під клієнта)	Unsupervised (кластеризація), але потребує ручної валідації та редагування таксономії	Supervised Learning (потребує ручного створення та навчання класифікаторів)	Rule-based / Keyword extraction (обмежена гнучкість)	Zero-shot LLM Discovery: повна автоматизація без ручного втручання
Контекстне вікно та Big Data	Обмежене, аналіз часто базується на окремих реченнях або фразах	Обмежене, фокус на частотності тем, лінійна залежність вартості від обсягу	Обмежене, класифікація окремих текстів, складно для довгих історій	Обмежене внутрішніми даними платформи	Необмежене (RAG + Long-context 1M+ tokens), аналіз історичних трендів та великих масивів ¹
Час налаштування (Time-to-Value)	Тижні (потребує інтеграції та налаштування моделей консультантами)	Дні/Тижні (потребує "навчання" системи на даних клієнта)	Тижні (користувач сам має тренувати моделі)	Миттєво (але тільки для даних, зібраних через Yotpo)	Миттєво (Plug & Play, перший звіт через 5 хв після підключення)
Глибина семантичного аналізу	Висока (Aspect-Based Sentiment Analysis), але складна в управлінні	Висока, детальна ієрархія тем, але складна для неспеціалістів	Середня/Висока (залежить від якості навчальної вибірки користувача)	Базова (Sentiment + Keywords), фокус на конверсію, а не аналітику	Висока (LLM розуміє нюанси, сарказм, імпліцитні аспекти) [14]

Кінець табл. 3.1

<i>Техніко-економічна характеристика</i>	<i>Chattermill</i>	<i>Thematic</i>	<i>MonkeyLearn</i>	<i>Yotpo (Analytics)</i>	<i>Наш проект (ReviewMind AI)</i>
<i>Генерація інсайтів</i>	"Що сталося" + "Вплив на NPS" (Impact Analysis)	"Чому це відбувається" (кореляції тем)	Тільки "Що" (теги та категорії)	Тренди настроїв, вплив на продажі	Prescriptive Analytics: "Що сталося, Чому, і Що робити" (Actionable Recommendations)
<i>Джерела даних</i>	Агностичний (будь-які джерела через API)	Агностичний (фокус на опитуваннях та тикетах)	Агностичний (CSV, API, Zapier)	Закрита екосистема (переважно дані Yotpo)	Мультиканальний (Native Connectors: Amazon, Shopify)

Chattermill є лідером у сегменті Enterprise завдяки потужній функції Impact Analysis, яка корелює теми відгуків з метриками NPS/CSAT, проте його ціна та складність впровадження роблять його недоступним для масового ринку [61].

Thematic пропонує якісний індуктивний аналіз, але позиціонується як преміальний інструмент для великих команд дослідників (Insights Teams) з бюджетом від \$25,000 на рік [57].

MonkeyLearn надає інструментарій (конструктор), а не готове рішення, що вимагає від клієнта значних зусиль на створення та підтримку моделей класифікації [63].

Yotpo зосереджений на зборі відгуків для маркетингу (UGC), а аналітика є вторинною функцією, доступною лише на найдорожчих тарифах [59].

Наш проект заповнює нішу між функціональністю дорогих Enterprise-рішень та доступністю простих інструментів, пропонуючи

унікальну комбінацію автоматизації (через LLM) та глибини аналізу (через необмежений контекст).

Технологічний аудит визначає оптимальний технологічний стек для реалізації заявлених функцій, враховуючи необхідність обробки великих даних, зниження операційних витрат та забезпечення масштабованості (табл. 3.2).

Таблиця 3.2 – Технологічна здійсненність ідеї проекту

<i>Технологія реалізації</i>	<i>Доступність</i>	<i>Придатність</i>	<i>Обрана</i>	<i>Обґрунтування</i>
LLM Orchestration (LangChain / LlamaIndex)	Відкритий код	Висока, індустріальний стандарт	Так	Дозволяє ефективно керувати ланцюжками запитів до LLM, інтегрувати зовнішні джерела даних та реалізувати логіку RAG [63].
Vector Database (Pinecone / Milvus / pgvector)	Відкритий код / SaaS	Висока, критична для семантичного пошуку	Так	Забезпечує швидкий пошук схожих відгуків для формування контексту.
Long-Context LLMs (Gemini 1.5 Pro / Claude 3 Opus)	Комерційний API	Висока, унікальна перевага	Так	Використання великих моделей дозволяє аналізувати тисячі відгуків одночасно для виявлення глобальних зв'язків [47].
Cost-Efficient Models (GPT-4o-mini / Llama 3)	API / Open Weights	Висока, для рутинних задач	Так	Використання дешевих моделей для попередньої обробки знижує собівартість (COGS) у 10-20 разів порівняно з флагманськими моделями [65].

Кінець табл. 3.2

<i>Технологія реалізації</i>	<i>Доступність</i>	<i>Придатність</i>	<i>Обрана</i>	<i>Обґрунтування</i>
Unsupervised Clustering (BERTopic / HDBSCAN)	Відкритий код	Висока, для discovery	Так	Дозволяє групувати відгуки за темами без попереднього навчання, що є основою функції автоматичної генерації аспектів [66].
Backend (Python + FastAPI)	Відкритий код	Висока, стандарт для AI-проектів	Так	Забезпечує високу продуктивність та легку інтеграцію з бібліотеками машинного навчання (PyTorch, Scikit-learn).
Frontend (React + Tailwind)	Відкритий код	Висока, сучасний стандарт	Так	Дозволяє створити швидкий та інтерактивний інтерфейс користувача для візуалізації складних аналітичних даних.
Infrastructure (AWS / Azure)	Комерційно доступна	Висока, надійність та масштабування	Так	Хмарна інфраструктура забезпечує масштабування обчислювальних потужностей (GPU) під час пікових навантажень [64].

3.2 Аналіз ринкових можливостей та конкурентного середовища

Глобальний ринок аналітики клієнтських даних (Customer Analytics Market) перебуває на стадії активного зростання. За оцінками Grand View Research, обсяг ринку досяг 14.57 мільярдів доларів у 2023 році та, за прогнозами, зросте до 48.63 мільярдів до 2030 року, демонструючи середньорічний темп зростання (CAGR) на рівні 19.2% [67].

Ще більш динамічним є зростання сегменту аналізу настроїв (Sentiment Analysis), який є ядром технології нашого проекту. Цей ринок, за оцінками, зросте з 3.9 мільярдів доларів у 2024 році до 17 мільярдів у 2030 році з вражаючим CAGR 27.7% [68]. Таке стрімке зростання зумовлене потребою бізнесів у реальному часі розуміти емоції клієнтів у цифровому просторі.

Ринок обробки природної мови (NLP) в сфері електронної комерції та рітейлу також показує значний потенціал, прогнозуючись досягти 439 мільярдів доларів до 2030 року при CAGR 38.7% [67]. Це свідчить про те, що технологічний фундамент проекту є надзвичайно актуальним і затребуваним.

Ринок характеризується змішаною структурою: присутні як великі універсальні платформи (Qualtrics, Medallia), так і спеціалізовані нішеві рішення. Проте, бар'єри входу залишаються середніми: хоча доступ до базових AI-технологій спростився, створення високоточної системи, що розуміє специфіку E-commerce без складного налаштування, вимагає глибокої технічної експертизи в RAG та архітектурі LLM. Попередня характеристика потенційного ринку наведена в табл. 3.3.

Таблиця 3.3 – Попередня характеристика потенційного ринку

<i>Показник</i>	<i>Характеристика</i>
Основні гравці	Yotpo (збір), Chattermill (аналітика), Thematic (теми), MonkeyLearn (ML), Medallia (CX), Qualtrics (XM)
Обсяг ринку (Customer Analytics) 2024	~\$17 млрд глобально [69]
Динаміка	+19.6% щорічно для загальної аналітики, +27.7% для аналізу настроїв [52]

Кінець табл 3.3

Бар'єри входу	Середні (технічна експертиза, доступ до даних, вартість обчислень)
Норма рентабельності	70-80% валової маржі (Gross Margin) для SaaS при ефективній оптимізації витрат на AI [70]

Цільові сегменти клієнтів для нашого стартапу можна розділити на три основні групи.

1. ***DTC Бренди (Mid-market E-commerce)***: компанії з річним оборотом від \$10 млн до \$100 млн. Вони отримують тисячі відгуків щомісяця через різні канали (власний сайт, Amazon, соцмережі), але не мають ресурсів для їх ручного аналізу. Для них критично важливо швидко виявляти дефекти продукту та причини повернень. Готовність платити висока, оскільки це напряму впливає на прибуток.
2. ***SaaS Компанії (Product & Customer Success Teams)***: команди, які потребують агрегації фідбеку з G2, Capterra, Intercom та Slack для пріоритезації продуктового беклогу. Їм потрібна не просто оцінка настрою, а виділення конкретних фіч, які потребують доопрацювання. Вони технічно грамотні і готові до впровадження нових інструментів.
3. ***Маркетингові та PR-агенції***: компанії, що обслуговують багато брендів і потребують інструментів для швидкого аудиту репутації клієнтів та порівняльного аналізу з конкурентами. Для них важлива швидкість генерації звітів та можливість "white-label" рішень.

У табл. 3.4 наведено аналіз конкуренції.

Основний виклик конкуренції долається через унікальну архітектуру. У той час як ***MonkeyLearn*** вимагає від користувача тренувати модель, а ***Thematic*** потребує часу на адаптацію таксономії, наше рішення на базі ***LLM***

Zero-shot Discovery починає працювати миттєво після підключення джерела даних.

Таблиця 3.4 – Аналіз конкуренції за моделлю М. Портера

<i>Сила</i>	<i>Характеристика</i>	<i>Інтенсивність</i>	<i>Вплив на проект</i>
Прямі конкуренти	Chattermill, Thematic, Yotpo (Reviews Atlas)	Висока	Необхідність чіткої диференціації через "автоматизацію без налаштувань" та ціну.
Потенційні конкуренти	Універсальні LLM (ChatGPT, Claude), Платформи (Trustpilot, Gorgias)	Середня	Фокус на спеціалізованих UX/UI та інтеграціях.
Постачальники	Провайдери LLM (OpenAI, Anthropic, Google), Хмарні сервіси (AWS)	Помірна	Зниження ризиків margin compression.
Клієнти	E-commerce бренди, SaaS компанії	Висока	Необхідність пропонувати прозору модель ціноутворення та швидкий Time-to-Value.
Замінники	Ручний аналіз (Excel/Google Sheets), Фрілансери, Власні розробки	Середня	Акцент на економії часу (90% reduction) та повноті покриття даних.

Крім того, більшість конкурентів мають обмеження на довжину тексту або кількість відгуків для одночасного аналізу. Наша RAG-архітектура знімає ці обмеження, дозволяючи "спілкуватися" з усією історією відгуків (Big Data).

Критичні фактори успіху включають: точність автоматичного виявлення аспектів, щоб уникнути не релевантних категорій, швидкість обробки великих масивів даних та безшовність інтеграцій з основними платформами (Shopify, Zendesk).

В табл 3.5 наведений SWOT-аналіз проекту.

Таблиця 3.5 – SWOT-аналіз проекту

Сильні сторони (Strengths)	Слабкі сторони (Weaknesses)
Технологія авто-генерації аспектів (Zero-shot) без ручного налаштування.	Відсутність впізнаваного бренду та репутації на ринку.
Необмежене контекстне вікно (Big Data ready) завдяки RAG та Long-Context LLMs.	Залежність від сторонніх API (OpenAI, Anthropic) та їх стабільності.
Конкурентна ціна завдяки каскадній обробці та оптимізації токенів.	Обмежені маркетингові бюджети порівняно з гігантами (Qualtrics, Yotpo).
Легкість впровадження (Plug & Play інтеграції).	Потреба у постійній адаптації до змін у політиках платформ (API ліміти).
Можливості (Opportunities)	Загрози (Threats)
Стрімке зростання ринку NLP в E-commerce (+38% CAGR).	Агресивне зниження цін конкурентами або впровадження ними аналогічних AI-функцій.
Високий попит на AI-аналітику серед SMB та Mid-market, які ігноруються Enterprise-гравцями.	Регуляторні ризики (GDPR, EU AI Act) щодо обробки персональних даних у хмарі.
Глобальна експансія завдяки мультимовності сучасних LLM (вихід на ЄС, Латам).	Поява безкоштовних вбудованих аналітичних інструментів у Shopify або Amazon.
Партнерства з маркетинговими агенціями для масштабування продажів.	"Галюцинації" моделей, що можуть призвести до невірних інсайтів та втрати довіри.

Стратегія виходу на ринок передбачає використання моделі **Product-Led Growth (PLG)**. Ми пропонуватимемо безкоштовний "Health Check" аудит: користувач підключає джерело відгуків і отримує безкоштовний аналіз останніх 100 відгуків з виявленням трьох головних проблем продукту. Це демонструє цінність продукту миттєво (Time-to-Value < 5 хвилин) і створює віральний ефект. Для середнього та великого бізнесу передбачено прямі продажі (Outbound Sales) з фокусом на демонстрацію ROI (зменшення churn rate та повернень).

3.3 Розроблення маркетингової стратегії та бізнес-моделі

Маркетингова стратегія проекту "ReviewMind AI" базується на глибокому розумінні потреб обраних цільових сегментів та пропозиції чіткої, вимірюваної цінності. Враховуючи високу конкуренцію, стратегія фокусується на "демократизації" складної аналітики, роблячи інструменти рівня Enterprise доступними для ширшого кола бізнесів.

Аналіз показує, що оптимальною є стратегія концентрованого зростання з початковим фокусом на DTC брендах та SaaS компаніях, які мають найвищу "больову точку" – втрату клієнтів через невідомі причини (табл. 3.6).

Базова стратегія розвитку визначається як стратегія диференціації (табл. 3.7). На відміну від Yotpo, який конкурує екосистемою маркетингових інструментів (лояльність, SMS), або MonkeyLearn, який конкурує як інструмент для розробників, наше позиціонування базується на інтелекті та автоматизації. Ми продаємо не "інструмент для аналізу тексту", а "автоматичного аналітика, який працює 24/7".

Базова стратегія розвитку визначається як стратегія диференціації (табл. 3.7). На відміну від Yotpo, який конкурує екосистемою маркетингових інструментів (лояльність, SMS), або MonkeyLearn, який конкурує як інструмент для розробників, наше позиціонування базується на інтелекті та автоматизації. Ми продаємо не "інструмент для аналізу тексту", а "автоматичного аналітика, який працює 24/7".

Конкурентна поведінка на перших етапах орієнтована на партизанський маркетинг у спільнотах E-commerce та SaaS, демонстрацію кейсів, де наш інструмент знайшов критичні проблеми, пропущені іншими.

Таблиця 3.6 – Вибір цільових груп потенційних споживачів

Цільовий сегмент	Характеристика	Готовність споживачів	Орієнтовний попит	Інтенсивність конкуренції	Простота входу
DTC Бренди (Mid-market)	Річний оборот \$10M-\$100 M. Отримують 500+ відгуків/міс. Висока чутливість до репутації та повернень.	Висока (шукають інструменти для зменшення повернень та покращення рейтингу)	Високий (типовий біль – ручна обробка CSV файлів або ігнорування тексту)	Висока (Yotpo, Gorgias), але конкуренти фокусуються на зборі/підтримці, а не глибинній аналітиці	Висока (через екосистему Shopify та BigCommerce)
SaaS Стартапи & Scaleups	Продуктові команди (Product Managers). Активно працюють над Product-Market Fit та зменшенням churn.	Висока (потребують агрегації фідбеку з багатьох каналів: G2, Intercom, Sales calls)	Середній (вимоги більш специфічні та технічні)	Середня (Chattermill, Dovetail, Thematic)	Середня (вимагає інтеграцій з CRM та Support системами)
Маркетингові Агенції	Обслуговують десятки клієнтів. Потрібна швидка генерація звітів для демонстрації результатів.	Середня (часто покладаються на ручну працю джуніорів)	Високий (потреба в автоматизації рутинної звітності)	Низька (мало спеціалізованих інструментів для агенцій)	Висока (прямі партнерства)

Стратегія позиціонування: "ReviewMind AI – єдина платформа, яка читає, розуміє та систематизує 100% відгуків ваших клієнтів, перетворюючи хаос думок на чіткий план дій для зростання".

Бізнес-модель базується на класичному B2B SaaS підході з багаторівневою підпискою (табл. 3.8). Ціноутворення прив'язане до обсягу оброблюваних даних (кількість відгуків або джерел), що безпосередньо корелює з цінністю для клієнта та витратами компанії (COGS).

Таблиця 3.7 – Базова стратегія розвитку

<i>Альтернатива стратегії</i>	<i>Обґрунтування вибору</i>
Стратегія диференціації	Унікальна технологія авто-генерації аспектів та "необмежений контекст" створюють чітку відмінність від конкурентів, які вимагають ручного налаштування або мають ліміти. Це дозволяє обґрунтувати цінність навіть при наявності дешевших альтернатив.
Стратегія нішевої спеціалізації	Фокус на E-commerce та SaaS дозволяє розробити специфічні інтеграції (Shopify, Zendesk) та шаблони звітів, які вирішують конкретні проблеми цих індустрій краще за універсальні інструменти (наприклад, аналіз причин повернень).
Географічна стратегія	"Global from Day One". Пріоритет – англомовні ринки (США, UK, Канада, Австралія) через високу платоспроможність та розвинену культуру E-commerce. Мультимовність LLM дозволяє легко масштабуватися на Європу без змін у продукті.

Таблиця 3.8 – Визначення базової бізнес-моделі

<i>Компонент</i>	<i>Опис</i>
Ціннісна пропозиція	Автоматичний аналіз 100% відгуків (Big Data) без складних налаштувань. Виявлення прихованих проблем якості ("Why churn?"). Економія до 90% часу аналітиків/менеджерів. Actionable Insights для покращення продукту.
Цільові сегменти	DTC бренди (Mid-market), SaaS продуктові команди, Маркетингові агенції.

Кінець табл. 3.8

<i>Компонент</i>	<i>Опис</i>
Канали збуту	Shopify App Store (для E-com), Direct Sales (LinkedIn Outreach, Demo), Content Marketing (SEO, Case Studies), Партнерська мережа.
Відносини з клієнтами	Self-service (для Starter/Pro), Dedicated Success Manager (для Growth/Enterprise), Community support.
Потоки доходів	Tiered Subscription: Starter (\$199/mic), Pro (\$499/mic), Growth (\$999/mic). Add-ons: Додаткові джерела даних, історичний аналіз архівів (one-time fee).
Ключові ресурси	Технологічна платформа (RAG, LLM integration), R&D команда, Sales & Marketing команда, Хмарна інфраструктура.
Ключові активності	Розробка продукту (Product Dev), Залучення клієнтів (Customer Acquisition), Підтримка та Success, Оптимізація AI-моделей.
Ключові партнери	Платформи E-commerce (Shopify, BigCommerce), Хмарні провайдери (AWS, Azure), LLM провайдери (OpenAI, Anthropic), Інтегратори.
Структура витрат	R&D (персонал), COGS (API витрати, сервери), Sales & Marketing (CAC), G&A (адміністративні).

Модель монетизації враховує специфіку витрат на AI. Оскільки кожна обробка відгуку коштує грошей (токени), безлімітні плани є ризикованими. Тому вводяться ліміти на кількість відгуків у кожному тарифі, що дозволяє контролювати маржинальність.

Для залучення клієнтів використовується комбінація вхідного (Inbound) та вихідного (Outbound) маркетингу (табл. 3.9).

Таблиця 3.9 – Концепція маркетингових комунікацій

<i>Цільова група</i>	<i>Канали комунікації</i>	<i>Ключові повідомлення</i>	<i>Бюджет (розподіл)</i>
DTC Бренди	Shopify App Store Ads, Instagram/LinkedIn Ads, Influencer Marketing (E-com експерти), SEO	"Дізнайтесь, чому повертають ваші товари, за 5 хвилин". "Аналітика рівня Amazon для вашого бренду". "Перетворіть негативні відгуки на лояльність".	50%
Product Managers (SaaS)	Content Marketing (Blog, Whitepapers, Cases), Product Hunt, Reddit/Hacker News, Спільноти (Slack, Discord)	"Перестаньте тегати тікети вручну. Нехай AI буде ваш роадмап". "Зрозумійте справжні причини Churn". "Голос клієнта без шуму".	30%
Агенції	Direct Outreach (Email, LinkedIn), Webinars, Partner Programs, Галузеві конференції	"Звіти для клієнтів за один клік". "Збільште ROI ваших послуг та LTV клієнтів". "Додайте AI-аналітику до вашого портфоліо".	20%

Стратегія ціноутворення використовує психологічний підхід "Good-Better-Best" [72].

1. Starter (\$199/міс): до 2,000 відгуків/міс. Включає: Аналіз настроїв, Авто-теми, Дашборд. Орієнтовано на малі бренди, яким потрібен базовий огляд.
2. Pro (\$499/міс): до 10,000 відгуків/міс. Включає: Starter + Історичний аналіз, Інтеграції (Zendesk, Slack), Розширений контекст, Експорт даних. "Sweet spot" для середніх компаній.
3. Growth (\$999+/міс): 50,000+ відгуків/міс. Включає: Pro + API доступ, Кастомні звіти, Пріоритетна підтримка, SSO. Для великих гравців та агенцій.

Графік виходу на ринок передбачає три фази:

1. MVP & Beta (місяці 1-3): запуск закритої бети для 10-20 дружніх компаній. Відлагодження RAG-пайплайну та точності моделей.
2. Launch & PLG (місяці 4-9): публічний запуск на Product Hunt, публікація додатку в Shopify App Store. Фокус на самообслуговуванні та зборі відгуків. Ціль: 50 платних клієнтів.
3. Scale & Outbound (місяці 10+): формування Sales-команди для роботи з Mid-market та Enterprise лідами. Розширення інтеграцій. Ціль: 250+ клієнтів.

Стратегія успішності клієнтів (Customer Success) є критичною для утримання (Retention) та розширення (Expansion) в B2B SaaS. Ми плануємо впровадити проактивний моніторинг використання продукту: якщо клієнт не заходить в систему або не генерує звіти, менеджер зв'язується з пропозицією допомоги. Також важливою є освітня складова – вебінари та кейси про те, як інтерпретувати інсайти та перетворювати їх на бізнес-рішення.

3.4 Фінансово-економічне обґрунтування проекту

Фінансово-економічне обґрунтування проекту базується на реалістичних оцінках витрат, притаманних AI-стартапам, та прогнозах доходів, що відповідають бенчмаркам SaaS-індустрії. Ключовою особливістю є врахування змінних витрат на використання LLM (Inference costs) у структурі собівартості.

Початкові інвестиційні потреби охоплюють період розробки MVP (6 місяців) та виходу на ринок (табл. 3.10). Основна стаття витрат – це команда розробників (R&D). Згідно з даними ринку праці, зарплата кваліфікованого AI/ML інженера в Україні становить близько \$3,500 - \$4,500 на місяць (\$40,800 - \$54,000 на рік) [74], що є значно економічнішим порівняно з США (\$120,000 - \$200,000) [76].

Таблиця 3.10 – Початкові інвестиційні потреби

Категорія витрат	Сума (USD)	Обґрунтування
Product Development (R&D, 6 міс)	\$240,000	Команда: 2 Senior AI Engineers (Backend/ML), 1 Full-stack Developer, 1 Product Manager/Designer. Розрахунок базується на конкурентних зарплатах для залучення талантів [74].
Хмарна інфраструктура та API (Setup)	\$20,000	Налаштування AWS/Azure, кредити на OpenAI/Anthropic API для тестування та валідації гіпотез, підписка на Vector DB (Pinecone/Milvus) [77].
Юридичне оформлення та IP	\$15,000	Реєстрація компанії (Delaware C-Corp), юридичний супровід, захист інтелектуальної власності, GDPR комплаєнс.

Кінець табл 3.10

Початковий маркетинг (Go-to-Market)	\$40,000	Розробка брендингу, веб-сайту, створення контенту, лістинг у Shopify App Store, початковий рекламний бюджет для трафіку.
Офіс (Remote-first) та ПЗ	\$10,000	Ліцензії на ПЗ (Jira, Slack, GitHub, Notion), обладнання (ноутбуки) для команди.
Operational Buffer (Резерв)	\$25,000	Резервний фонд на непередбачувані витрати (близько 7% від бюджету).
Загальні початкові потреби	\$350,000	Необхідний обсяг фінансування для Pre-seed раунду або bootstrapping до запуску MVP.

Операційні витрати після запуску включатимуть змінні витрати на інфраструктуру (COGS - Cost of Goods Sold), які для AI-проектів можуть бути суттєвими. Використання моделі GPT-4o-mini (\$0.15/1M вхідних токенів) дозволяє тримати ці витрати під контролем [78]. Для обробки 10,000 відгуків (приблизно 1.5 млн токенів) витрати на API складуть менше \$1, що забезпечує високу валову маржу. Однак, використання потужних моделей (Gemini 1.5 Pro) для складних запитів підвищує середню собівартість. Прогнозується, що COGS становитиме близько 20-25% від доходу, що відповідає здоровим показникам SaaS (табл 3.11).

Прогноз доходів базується на поступовому нарощуванні клієнтської бази. Середній дохід на користувача (ARPU) планується на рівні \$300-\$550/міс, що є конкурентним порівняно з Chattermill та Thematic, але вищим за прості інструменти. Вартість залучення клієнта (CAC) для B2B SaaS у сегменті аналітики варіюється від \$200 до \$700+ [80]. Ми закладаємо

консервативний САС, який буде знижуватися з часом завдяки PLG та впізнаваності бренду (табл. 3.12).

Таблиця 3.11 – Прогнозовані операційні витрати (USD на рік)

<i>Категорія</i>	<i>Рік 1</i>	<i>Рік 2</i>	<i>Рік 3</i>
R&D (Engineering & Product)	\$500,000	\$900,000	\$1,500,000
Sales & Marketing (S&M)	\$150,000	\$450,000	\$1,200,000
Infrastructure & AI APIs (COGS)	\$30,000	\$120,000	\$350,000
Customer Success (CS)	\$0 (Founder led)	\$100,000	\$300,000
G&A (Admin, Legal, Finance)	\$50,000	\$100,000	\$250,000
Загальні операційні витрати	\$730,000	\$1,670,000	\$3,600,000

Таблиця 3.12 – Прогноз доходів та ключових метрик

<i>Метрика</i>	<i>Рік 1</i>	<i>Рік 2</i>	<i>Рік 3</i>
Кількість платних клієнтів (кінець року)	50	250	800
ARPU (Середній дохід на клієнта/міс)	\$300	\$400	\$550
MRR (Щомісячний регулярний дохід)	\$15,000	\$100,000	\$440,000

Кінець табл 3.12

ARR (Щорічний регулярний дохід)	\$180,000	\$1,200,000	\$5,280,000
Revenue Growth Rate	-	566%	340%
CAC (Вартість залучення клієнта)	\$3,000	\$1,800	\$1,500
LTV (Позиттєва цінність клієнта)	\$5,400	\$9,600	\$16,500
LTV/CAC Ratio	1.8x	5.3x	11.0x

Примітка: LTV розраховано виходячи з очікуваного рівня відтоку (Churn Rate) 3-5% на місяць для малого бізнесу та <1% для Enterprise, що дає середній час життя клієнта 18-30 місяців [82]. Зростання ARPU пояснюється переходом клієнтів на вищі тарифні плани (Pro/Growth) та розширенням функціоналу (табл 3.13).

Таблиця 3.13 – Прогноз cash flow та прибутковості (USD)

Показник	Рік 1	Рік 2	Рік 3
Загальний дохід (Revenue)	\$180,000	\$1,200,000	\$5,280,000
Собівартість (COGS ~20%)	\$36,000	\$240,000	\$1,056,000
Валовий прибуток (Gross Profit)	\$144,000	\$960,000	\$4,224,000
Валова маржа (Gross Margin)	80%	80%	80%

Кінець табл. 3.13

Валова маржа (Gross Margin)	80%	80%	80%
Операційні витрати (без COGS)	\$694,000	\$1,430,000	\$2,544,000
EBITDA	-\$550,000	-\$470,000	+\$1,680,000
Потреба у фінансуванні (cumulative)	~\$900,000	~\$1,400,000	Self-sustaining

Проект демонструє класичну "J-curve" для венчурних стартапів. Перші два роки є інвестиційними (burn period), де витрати перевищують доходи через активну розробку та залучення клієнтів. Вихід на точку беззбитковості (EBITDA Breakeven) прогнозується на третій рік при досягненні ARR ~\$2-3 млн.

Загальна потреба у зовнішньому фінансуванні до виходу на самоокупність оцінюється в \$1.5 - \$2.0 млн. Це може бути покрито через раунд Seed (\$500k-\$750k) та подальший Series A.

Інвестиційна привабливість підтверджується високим потенційним мультиплікатором оцінки. SaaS компанії з темпами зростання >100% (T2D3) часто оцінюються в 8-15x від ARR. При досягненні ARR \$5.28 млн на 3-й рік, оцінка компанії може скласти \$40 - \$80 млн, що забезпечить значний прибуток (ROI) для інвесторів ранніх стадій.

Ризики включають потенційне зростання цін на AI API або посилення конкуренції ("price war"). Однак, тенденція до здешевлення токенів (наприклад, GPT-4o-mini дешевший за GPT-3.5 Turbo на 60% [79]) та можливість використання open-source моделей на власних серверах (Llama 3) є ефективними стратегіями хеджування цих ризиків.

Висновки до розділу 3

У четвертому розділі проведено комплексний аналіз комерціалізації проекту "ReviewMind AI" – інноваційної платформи для аналізу відгуків.

1. **Продукт:** розроблено концепцію системи, що вирішує проблему аналізу Big Data відгуків через унікальну комбінацію "необмеженого контексту" (RAG + Long-context LLM) та "автоматичної генерації аспектів" (Zero-shot). Це забезпечує значну технологічну перевагу над традиційними гравцями (MonkeyLearn, Yotpo), які покладаються на ручні налаштування або прості алгоритми.
2. **Ринок:** проект виходить на ринок Customer Analytics (\$17 млрд) та Sentiment Analysis (\$3.9 млрд), що зростають з темпами 19-27% щорічно. Визначено вільну нішу в сегменті Mid-market (DTC та SaaS), який недостатньо обслуговується дорогими Enterprise-платформами (Chattermill, Thematic).
3. **Конкуренція:** стратегія диференціації дозволяє уникнути прямого зіткнення з гігантами завдяки фокусу на простоті ("Actionable insights in minutes") та доступній ціні (\$199-\$999/міс), що стало можливим завдяки оптимізації витрат на AI.
4. **Фінанси:** бізнес-модель SaaS з високою валовою маржею (80%) є життєздатною. Проект потребує початкових інвестицій ~\$350,000 для запуску MVP та загалом ~\$1.5-2.0 млн для масштабування, з прогнозованим виходом на прибутковість на 3-й рік та досягненням ARR \$5.28 млн.

Таким чином, проект має високий потенціал для реалізації, чітку ринкову стратегію та привабливі фінансові показники для інвесторів.

ВИСНОВКИ

У дипломній роботі вирішено актуальне науково-прикладне завдання розробки методології та архітектури системи для автоматизованої трансформації відгуків користувачів у кількісне аспектно-орієнтоване резюме. Проведене дослідження траєкторії розвитку методів обробки природної мови показало чітку міграцію від загального узагальнення тексту до інтелектуального аналізу думок (Opinion Mining). Встановлено, що існуючі підходи часто є неефективними для обробки E-commerce відгуків, оскільки не враховують повною мірою їхні специфічні властивості: суб'єктивність, високий рівень зашумленості, часову динаміку та напівструктурований характер. Визначено, що найбільш перспективним напрямком є створення систем, які поєднують LLM з аспектно-орієнтованим підходом.

У роботі запропоновано та формалізовано новий стандарт кількісного аспектно-орієнтованого резюмування (Quantified Aspect-Based Summary – QABS). Стандарт визначено за сімома ключовими критеріями та представлено уніфікованим шаблоном у форматі JSON. Порівняльний аналіз із існуючими рішеннями провідних IT-компаній підтвердив перевагу QABS у структурованості та інформативності, що дозволяє перетворити неструктурований текст відгуків у дані, придатні для кількісного аналізу.

Розроблено модульну архітектуру системи, яка забезпечує реалізацію стандарту QABS. Ключовою особливістю розробки стало впровадження спеціальних метрик для оцінювання фактичної узгодженості (factuality) згенерованого тексту. Співставлення запропонованих метрик із традиційними методами оцінки (ROUGE, BERTScore, G-Eval, QAG) довело, що розроблений підхід дозволяє краще контролювати достовірність резюмування та мінімізувати галюцинації генеративних моделей.

Економічну доцільність розробки підтверджено комплексним аналізом стартап-проекту «ReviewMind AI». Продукт орієнтований на зростаючі ринки Customer Analytics (\$17 млрд) та Sentiment Analysis, займаючи вільну нішу в сегменті Mid-market (DTC та SaaS). Завдяки використанню технології RAG та Long-context LLM, система забезпечує конкурентну перевагу над традиційними гравцями (MonkeyLearn, Yotpo). Розрахована фінансова модель SaaS з валовою маржею 80% та прогнозованим виходом на прибутковість на 3-й рік свідчить про високу інвестиційну привабливість проекту.

Наукова новизна роботи полягає у впровадженні стандарту кількісного аспектно-орієнтованого резюмування, розробці архітектури системи для його реалізації та впровадженні метрик оцінювання фактичної узгодженості згенерованого тексту. Практична цінність результатів полягає у створенні готового інструментарію для продавців та покупців, що дозволяє швидко отримувати об'єктивну інформацію з великих масивів відгуків.

Результати дисертаційного дослідження апробовано на науковій конференції та опубліковано в матеріалах конференцій, що підтверджує їх наукову новизну та практичну значущість.

ПЕРЕЛІК ПОСИЛАНЬ

1. A Comprehensive Survey on Automatic Text Summarization with Exploration of LLM-Based Methods - arXiv, доступ отримано грудня 16, 2025, <https://arxiv.org/html/2403.02901v2>
2. Newegg Summarizes Customer Reviews Through AI - CMSWire, доступ отримано грудня 16, 2025, <https://www.cmswire.com/the-wire/newegg-summarizes-customer-reviews-through-ai/>
3. Extractive Opinion Summarization in Quantized Transformer Spaces - MIT Press Direct, доступ отримано грудня 16, 2025, https://direct.mit.edu/tacj/article/doi/10.1162/tacj_a_00366/98621/Extractive-Opinion-Summarization-in-Quantized
4. [2404.05243] Product Description and QA Assisted Self-Supervised Opinion Summarization, доступ отримано грудня 16, 2025, <https://arxiv.org/abs/2404.05243>
5. Findings of the Association for Computational Linguistics (2024) - ACL Anthology, доступ отримано грудня 16, 2025, <https://aclanthology.org/events/findings-2024/>
6. arXiv:2404.05243v1 [cs.CL] 8 Apr 2024 - CSE IITB, доступ отримано грудня 16, 2025, <https://www.cse.iitb.ac.in/~swaprava/papers/siledar24summarization.pdf>
7. FARSum: Feedback-aware abstractive product review ..., доступ отримано грудня 16, 2025, <https://www.amazon.science/publications/farsum-feedback-aware-abstractive-product-review-summarization>
8. FARSum: Feedback-Aware Abstractive Product ... - Amazon Science, доступ отримано грудня 16, 2025,

- <https://assets.amazon.science/9a/af/292c37b246edb20001f3e0a096b8/farsum-feedback-aware-abstractive-product-review-summarization.pdf>
9. FARSum: Feedback-Aware Abstractive Product Review Summarization - ResearchGate, доступ отримано грудня 16, 2025,
https://www.researchgate.net/publication/378199913_FARSum_Feedback-Aware_Abstractive_Product_Review_Summarization
 10. Beyond the Context Window: A New Approach to Summarizing Big ..., доступ отримано грудня 16, 2025,
<https://medium.com/flipkart-engineering/beyond-the-context-window-a-new-approach-to-summarizing-big-data-44b306a9608a>
 11. How does vector search compare to keyword search? - Milvus, доступ отримано грудня 16, 2025,
<https://milvus.io/ai-quick-reference/how-does-vector-search-compare-to-keyword-search>
 12. When Should You Use Full-Text Search vs. Vector Search? - Tiger Data, доступ отримано грудня 16, 2025,
<https://www.tigerdata.com/learn/full-text-search-vs-vector-search>
 13. [2509.26103] End-to-End Aspect-Guided Review Summarization at Scale - arXiv, доступ отримано грудня 16, 2025, <https://arxiv.org/abs/2509.26103>
 14. End-to-End Aspect-Guided Review Summarization at Scale - ACL Anthology, доступ отримано грудня 16, 2025,
<https://aclanthology.org/2025.emnlp-industry.31/>
 15. (PDF) Aspect-Based Opinion Summarization with Argumentation Schemes - ResearchGate, доступ отримано грудня 16, 2025,
https://www.researchgate.net/publication/392597807_Aspect-Based_Opinion_Summarization_with_Argumentation_Schemes
 16. Aspect-Based Opinion Summarization with Argumentation Schemes - arXiv, доступ отримано грудня 16, 2025,
<https://arxiv.org/html/2506.09917v2>

17. Newegg launches generative AI product review summary - Digital Commerce 360, доступ отримано грудня 16, 2025,
<https://www.digitalcommerce360.com/2023/08/24/newegg-launches-generative-ai-product-review-summary/>
18. Going beyond AI assistants: Examples from Amazon.com reinventing industries with generative AI | Artificial Intelligence, доступ отримано грудня 16, 2025,
<https://aws.amazon.com/blogs/machine-learning/going-beyond-ai-assistants-examples-from-amazon-com-reinventing-industries-with-generative-ai/>
19. A complete guide to Shopify Magic App Review Summaries - eesel AI, доступ отримано грудня 16, 2025,
<https://www.eesel.ai/blog/shopify-magic-app-review-summaries>
20. SLM vs LLM: The Key Differences - WEKA, доступ отримано грудня 16, 2025, <https://www.weka.io/learn/ai-ml/slm-vs-llm/>
21. Fine-Tuning Small Language Models to Optimize Code Review Accuracy | NVIDIA Technical Blog, доступ отримано грудня 16, 2025,
<https://developer.nvidia.com/blog/fine-tuning-small-language-models-to-optimize-code-review-accuracy/>
22. Beyond ROUGE: A Comprehensive Evaluation Metric for Abstractive Summarization Leveraging Similarity, Entailment, and Acceptability | International Journal on Artificial Intelligence Tools, доступ отримано грудня 16, 2025,
<https://www.worldscientific.com/doi/10.1142/S0218213024500179>
23. G-Eval Simply Explained: LLM-as-a-Judge for LLM Evaluation - Confident AI, доступ отримано грудня 16, 2025,
<https://www.confident-ai.com/blog/g-eval-the-definitive-guide>
24. FineSurE: Fine-grained Summarization Evaluation using LLMs | alphaXiv, доступ отримано грудня 16, 2025,
<https://www.alphaxiv.org/overview/2407.00908v3>

25. FineSurE: Fine-grained Summarization Evaluation using LLMs - arXiv, доступ отримано грудня 16, 2025, <https://arxiv.org/html/2407.00908v1>
26. FineSurE: Fine-grained Summarization Evaluation using LLMs - ACL Anthology, доступ отримано грудня 16, 2025, <https://aclanthology.org/2024.acl-long.51.pdf>
27. Entity hallucination index for text summarization models - Emerald Publishing, доступ отримано грудня 16, 2025, <https://www.emerald.com/aci/article/doi/10.1108/ACI-09-2024-0371/1301666/Entity-hallucination-index-for-text-summarization>
28. FaithBench: A Diverse Hallucination Benchmark for Summarization by Modern LLMs - arXiv, доступ отримано грудня 16, 2025, <https://arxiv.org/html/2410.13210v1>
29. Polarity Calibration for Opinion Summarization - arXiv, доступ отримано грудня 16, 2025, <https://arxiv.org/html/2404.01706v1>
30. Abstractive opinion summarization system (SelSum) and the largest dataset of Amazon product summaries (AmaSum). EMNLP 2021 conference paper. - GitHub, доступ отримано грудня 16, 2025, <https://github.com/abrazinskas/SelSum>
31. FaithBench: A Diverse Hallucination Benchmark for Summarization by Modern LLMs - ACL Anthology, доступ отримано грудня 16, 2025, <https://aclanthology.org/2025.naacl-short.38.pdf>
32. stangelid/qt - GitHub, доступ отримано грудня 12, 2025, <https://github.com/stangelid/qt>
33. Extractive Opinion Summarization in Quantized Transformer Spaces - MIT Press Direct, доступ отримано грудня 12, 2025, https://direct.mit.edu/tac/article/doi/10.1162/tac_l_a_00366/98621/Extractive-Opinion-Summarization-in-Quantized
34. Unsupervised Opinion Summarization Using Approximate Geodesics - ACL Anthology, доступ отримано грудня 12, 2025,

- <https://aclanthology.org/2023.newsum-1.11.pdf>
35. [2305.11603] Attributable and Scalable Opinion Summarization - arXiv, доступ отримано грудня 12, 2025, <https://arxiv.org/abs/2305.11603>
 36. Simple Yet Effective Synthetic Dataset Construction for Unsupervised Opinion Summarization - Amazon Science, доступ отримано грудня 12, 2025, <https://assets.amazon.science/9a/0f/73fbdef941f0813fc5d6b0dc74b4/simple-yet-effective-synthetic-dataset-construction-for-unsupervised-opinion-summarization.pdf>
 37. [2303.11660] Simple Yet Effective Synthetic Dataset Construction for Unsupervised Opinion Summarization - arXiv, доступ отримано грудня 12, 2025, <https://arxiv.org/abs/2303.11660>
 38. Summarizing Opinions: Aspect Extraction Meets Sentiment Prediction and They Are Both Weakly Supervised - Stefanos Angelidis, доступ отримано грудня 12, 2025, <https://stangelid.github.io/emnlp18oposum.pdf>
 39. arXiv:2109.04325v1 [cs.CL] 9 Sep 2021, доступ отримано грудня 12, 2025, <https://arxiv.org/pdf/2109.04325>
 40. OpenAsp: A Benchmark for Multi-document Open Aspect-based Summarization - arXiv, доступ отримано грудня 12, 2025, <https://arxiv.org/html/2312.04440v1>
 41. End-to-End Aspect-Guided Review Summarization at Scale - ACL Anthology, доступ отримано грудня 16, 2025, <https://aclanthology.org/2025.emnlp-industry.31>
 42. What are Review Highlights? https://biz.yelp.com/support-center/Yelp_Business_Page/Getting_Started/What-are-Review-Highlights/en-US
 43. Explore comment topics <https://support.google.com/youtube/answer/14239105?hl=en>
 44. TripAdvisor

https://www.tripadvisor.com/Hotel_Review-g27501570-d7180030-Reviews-La_Siesta_Classic_Ma_May_Hanoi_Hotel-Hoan_Kiem_Hanoi.html

- | | Reddit | Answers |
|-----|--|---|
| 45. | Introducing | https://redditinc.com/news/introducing-reddit-answers |
| 46. | An LLM-Based Approach to Review Summarization on the App Store | https://machinelearning.apple.com/research/app-store-review |
| 47. | Infinite Retrieval: Attention Enhanced LLMs in Long-Context Processing - arXiv, доступ отримано грудня 19, 2025, | https://arxiv.org/html/2502.12962v1 |
| 48. | Infinite Context Length in LLMs – The Next Big Advantage in AI by Aloy Banerjee Medium, доступ отримано грудня 19, 2025, | https://medium.com/@aloy.banerjee30/infinite-context-length-in-llms-the-next-big-advantage-in-ai-2550e9e6ce9b |
| 49. | A Hybrid Approach To Aspect Based Sentiment Analysis Using Transfer Learning - arXiv, доступ отримано грудня 19, 2025, | https://arxiv.org/html/2403.17254v1 |
| 50. | Case Study 2: An Unsupervised Neural Attention Model for Aspect Extraction - Medium, доступ отримано грудня 19, 2025, | https://medium.com/data-science/case-study-2-an-unsupervised-neural-attention-model-for-aspect-extraction-1c2c97b1380a |
| 51. | LLM Cost & Performance Analysis: Maximize Value - i10X, доступ отримано грудня 19, 2025, | https://i10x.ai/blog/llm-cost-and-performance-analysis |
| 52. | Global Customer Analytics Market Size & Outlook, 2025-2030, доступ отримано грудня 19, 2025, | https://www.grandviewresearch.com/horizon/outlook/customer-analytics-market-size/global |
| 53. | Compare Chattermill vs. Thematic G2, доступ отримано грудня 19, 2025, | https://www.g2.com/compare/chattermill-vs-thematic |

54. Chattermill vs. Thematic | Comparison, доступ отримано грудня 19, 2025, <https://chattermill.com/compare/chattermill-vs-thematic>
55. 8 Best Text Analytics Software (2025) - Thematic, доступ отримано грудня 19, 2025, <https://getthematic.com/insights/the-best-text-analytics-software>
56. Flexible Plans and Pricing for AI Feedback Analytics | Thematic, доступ отримано грудня 19, 2025, <https://getthematic.com/product/plan-and-pricing>
57. MonkeyLearn 2025 Pricing, Features, Reviews & Alternatives - GetApp, доступ отримано грудня 19, 2025, <https://www.getapp.com/business-intelligence-analytics-software/a/monkeylearn/>
58. 15 Sentiment Analysis Tools Ranked by Features & Cost, доступ отримано грудня 19, 2025, <https://www.kapiche.com/blog/sentiment-analysis-tools>
59. Customer Reviews 101: Your Guide To Impact & Strategies - Yotpo, доступ отримано грудня 19, 2025, <https://www.yotpo.com/blog/customer-reviews-101/>
60. Aspect-Sentiment Analysis for Scalable User Research | by Alex Crowley - Medium, доступ отримано грудня 19, 2025, <https://medium.com/@crowleyalex783/aspect-sentiment-analysis-for-scalable-user-research-bf3779c3d9ac>
61. Chattermill vs Unwrap.ai: Which Feedback Analytics Tool Wins? - SentiSum, доступ отримано грудня 19, 2025, <https://www.sentisum.com/library/chattermill-vs-unwrap>
62. Top 3 MonkeyLearn Alternatives & Competitors in 2025 - SentiSum, доступ отримано грудня 19, 2025, <https://www.sentisum.com/library/monkeylearn-alternatives-competitors>
63. Modern Tech Stack for Retrieval Augmented Generation (RAG) - Firecrawl, доступ отримано грудня 19, 2025,

- <https://www.firecrawl.dev/blog/modern-rag-tech-stack>
64. Build a RAG data ingestion pipeline for large-scale ML workloads - AWS, доступ отримано грудня 19, 2025,
<https://aws.amazon.com/blogs/big-data/build-a-rag-data-ingestion-pipeline-for-large-scale-ml-workloads/>
 65. GPT-4o Mini Pricing Calculator - NeuralTalk AI, доступ отримано грудня 19, 2025, <https://neuraltalk.ai/calculators/gpt-4o-mini>
 66. An unsupervised aspect extraction strategy for monitoring real-time reviews stream, доступ отримано грудня 19, 2025,
https://www.researchgate.net/publication/325064348_An_unsupervised_aspect_extraction_strategy_for_monitoring_real-time_reviews_stream
 67. Customer Analytics Market Size And Share Report, 2030, доступ отримано грудня 19, 2025,
<https://www.grandviewresearch.com/industry-analysis/customer-analytics-market-report>
 68. Sentiment Analysis - Social Media Analytics Market Statistics, доступ отримано грудня 19, 2025,
<https://www.grandviewresearch.com/horizon/statistics/social-media-analytics-market/technology/sentiment-analysis/global>
 69. Global Customer Analytics Market | 2019 – 2030 - Ken Research, доступ отримано грудня 19, 2025,
<https://www.kenresearch.com/global-customer-analytics-market>
 70. How to determine cost of goods sold (COGS) for your SaaS business - Drivetrain, доступ отримано грудня 19, 2025,
<https://www.drivetrain.ai/post/cost-of-goods-sold-cogs>
 71. LLM Cost Comparison 2025: A Deep Dive into Managing Your AI Budget, доступ отримано грудня 19, 2025,
<https://skywork.ai/skypage/en/LLM%20Cost%20Comparison%202025%3A%20A%20Deep%20Dive%20into%20Managing%20Your%20AI%20Budget>

[t/1975592241004736512](https://www.recurly.com/blog/how-to-create-an-effective-saas-pricing-strategy/)

72. SaaS Pricing Models & Strategy: A Guide - Recurly, доступ отримано грудня 19, 2025,
<https://recurly.com/blog/how-to-create-an-effective-saas-pricing-strategy/>
73. Your Ultimate Guide to SaaS Pricing Models - Revenera, доступ отримано грудня 19, 2025,
<https://www.revenera.com/blog/software-monetization/saas-pricing-models-guide/>
74. AI Engineer Salaries in 2025: Comprehensive Guide - Qubit Labs, доступ отримано грудня 19, 2025, <https://qubit-labs.com/ai-engineer-salary-guide/>
75. AI Engineer Salary 2025 : Overview, Trends, Optimization - Mobilunity, доступ отримано грудня 19, 2025,
<https://mobilunity.com/blog/ai-engineer-salary/>
76. Seed Startup Salary & Equity Compensation 2025, доступ отримано грудня 19, 2025,
<https://topstartups.io/startup-salary-equity-database/?stage=Seed>
77. Cost to Host and Scale a Private LLM in 2025 – Pricing & Factors Explained - Aimprosoft, доступ отримано грудня 19, 2025,
<https://www.aimprosoft.com/blog/cost-to-host-private-llm-2025/>
78. Claude 3.5 vs GPT 4o: Which LLM Reigns Supreme? - PromptLayer Blog, доступ отримано грудня 19, 2025,
<https://blog.promptlayer.com/big-differences-claude-3-5-vs-gpt-4o/>
79. GPT-4o mini: advancing cost-efficient intelligence - OpenAI, доступ отримано грудня 19, 2025,
<https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/>
80. Average Customer Acquisition Cost: Benchmarks by Industry - Userpilot, доступ отримано грудня 19, 2025,
<https://userpilot.com/blog/average-customer-acquisition-cost/>
81. Proven Strategies to Lower B2B SaaS Customer Acquisition Costs in 2025 -

Т.А. Монрое, доступ отримано грудня 19, 2025,

<https://tamonroe.com/articles/how-to-reduce-saas-marketing-cac/>

82. B2B SaaS Benchmarks: A Complete Guide 2026 - Churnfree, доступ

отримано грудня 19, 2025,

<https://churnfree.com/blog/b2b-saas-churn-rate-benchmarks/>

ДОДАТОК А ПРИКЛАДИ QABS

1. Example 1: book - "A Study in Scarlet", Arthur Conan Doyle

Product reviews summary:

"A Study in Scarlet" is the first Sherlock Holmes novel by Arthur Conan Doyle, and opinions on its quality are quite varied among Redditors. Here's a summary of what people think:

Pros

- **Character Introduction:** Many readers appreciate how the book introduces the iconic characters of Sherlock Holmes and Dr. John Watson. ["The characters of Holmes and Watson live up to their iconic reputations. Every page with them is a joy."](#)
(from 320 users, 94% agree)
- **Writing Style:** The prose is often praised for being engaging and lively. ["The prose is fun and lively. Many other works from this era can feel like a chore to read, but this one felt very modern while also capturing the time period."](#)
(from 215 users, 82% agree)
- **Historical Significance:** As the first Sherlock Holmes story, it holds a special place in detective fiction. ["Conan Doyle was writing before there were any genre tropes. The rule that a good mystery story should 'play fair' with the reader was a rule that only developed in the course of his writing."](#)
(from 150 users, 96% agree)

Cons

- **Sudden Backstory:** The sudden shift to a long backstory in the middle of the book is a major point of criticism. ["It's because it's sudden and jarring. Holmes and Watson are suddenly not in the story any more and we're given no warning for this."](#)
(from 400 users, 91% agree)
- **Pacing Issues:** The backstory is often seen as dragging and unnecessarily long. ["The second part was too long in my opinion. I don't think it was necessary to talk so much about Ferrier's backstory."](#)
(from 280 users, 78% agree)

- **Anti-Mormon Sentiment:** The portrayal of Mormons in the backstory has been criticized for being based on stereotypes and yellow journalism. ["I could do a longer explanation, but a lot of it is because it is considered anti-Mormon."](#)
(from 210 users, 85% agree)

Recommendations

- **Skip the Backstory:** Some readers suggest skipping the backstory if you find it tedious. ["You can, technically skip it except for the last few pages when it is centered on Holmes again."](#)
(from 110 users, 45% agree)
- **Read Other Stories First:** For a better introduction to Sherlock Holmes, some recommend starting with the short stories or other novels. ["I always recommend readers just work through the short stories first as it gives them what they're expecting from the characters and more succinctly."](#)
(from 175 users, 65% agree)

2. Example 2: smartphone - iPhone 16

Product reviews summary:

The iPhone 16 has received a mixed bag of reviews from Redditors. Here's a succinct guide summarizing their opinions to help you decide if it's the right upgrade for you.

Pros:

- **Camera Improvements:** The iPhone 16 Pro includes new tone control in the camera, allowing users to dial back HDR processing. This has been a significant relief for many users who were frustrated with the aggressive HDR in previous models. ["New tone control in camera lets you dial back HDR processing"](#)
(from 215 users, 85% agree)
- ...

Cons

- **Pricing:** The iPhone 16 is seen as overpriced by some users, especially when compared to the older models that offer similar features at a lower cost. ["Its just that its overpriced, would be very popular at a 20% lower price."](#)
(from 420 users, 65% agree)
- ...

User Experiences

- **Upgrade from Older Models:** Users upgrading from older models like the iPhone 11 or 12 have found the iPhone 16 to be a significant improvement in terms of battery life, camera, and overall performance. ["Upgraded from a 12 mini to a 16pro and it's amazing. The battery is great and the camera is way better. Loving it so far."](#)
(from 125 users, 92% agree)
- ...

Conclusion

The iPhone 16 is a solid phone with several improvements over older models, especially in camera and battery life. However, the high price and underdeveloped AI features might be a deterrent for some. If you are upgrading from an iPhone 11 or older, the iPhone 16 is likely to be a worthwhile investment. For users with newer models, the decision might depend on how much you value the minor improvements and new features