НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
Приладобудівний факультет
Кафедра автоматизації та систем неруйнівного контролю

«На правах рукопису»
УДК _____

До захисту допущено:
Завідувач кафедри
_____ Юрій КИРИЧУК
«___»_____20__ р.

Магістерська дисертація

на здобуття ступеня магістра

за освітньо-професійною програмою «Комп'ютерно-інтегровані системи та технології в приладобудуванні»

зі спеціальності 174 «Автоматизація, комп'ютерно-інтегровані технології та робототехніка»

на тему: «Автоматизована система контролю якості комунікації між клієнтами та персоналом контакт-центру»

Виконав:
студент ІІ курсу, групи ПК-41мп
Журавель Микола Миколайович _____

Науковий керівник:
Професор, д.п.н., професор кафедри АСНК
Протасов Анатолій Георгійович _____

Консультант з стартап-проекту:
Завідувач кафедри економічної кібернетики, д.е.н, професор
Бояринова Катерина Олександрівна _____

Рецензент:
Доцент каф. ІВТ, к.т.н. доцент
Маркіна Ольга Миколаївна _____

Засвідчую, що у цій магістерській дисертації немає запозичень з праць інших авторів без відповідних посилань.
Студент _____

Київ – 2025 року

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»

Приладобудівний факультет

Кафедра автоматизації та систем неруйнівного контролю

Рівень вищої освіти – другий (магістерський)

Спеціальність – 174 «Автоматизація, комп'ютерно-інтегровані технології та робототехніка»

Освітньо-професійна програма «Комп'ютерно-інтегровані системи та технології в приладобудуванні»

ЗАТВЕРДЖУЮ

Завідувач кафедри

_____ Юрій КИРИЧУК

«___»_____20__ р.

ЗАВДАННЯ
на магістерську дисертацію студенту

Журавлю Миколі Миколайовичу

1. Тема дисертації: «Автоматизована система контролю якості комунікації між клієнтами та персоналом контакт-центру», науковий керівник дисертації професор, д.п.н., професор кафедри АСНК Протасов Анатолій Георгійович, затверджені наказом по університету від «____» _____ 202___ р. № _____
2. Термін подання студентом дисертації _____
3. Об’єкт дослідження: процес контролю якості комунікацій у контакт-центрах, що включає інформаційні, програмні та організаційні компоненти систем автоматизації, призначені для перевірки відповідності взаємодії між клієнтом і персоналом встановленим стандартам обслуговування.
4. Вихідні дані: діалоги контакт-центру, стандарти обслуговування, критерії оцінювання якості, експертні оцінки та програмно-технічні засоби для реалізації прототипу автоматизованої системи контролю якості.
5. Перелік завдань, які потрібно розробити:
    - Проаналізувати сучасні підходи до автоматизації контролю якості комунікацій у контакт-центрах та визначити їх основні обмеження.
    - Обґрунтувати концептуальні та теоретичні засади побудови автоматизованої системи контролю якості, включаючи вимоги до архітектури, інфраструктури та методів оцінювання.

- Розробити багаторівневу модель оцінювання комунікацій, що включає критерії, підкритерії та формалізовані індикатори порушень.
- Дослідити можливості застосування мовних моделей (LLM) для автоматизованого аналізу діалогів, виявлення помилок та формування доказових (evidence) фрагментів.
- Розробити, реалізувати та експериментально перевірити MVP автоматизованої системи контролю якості шляхом порівняння результатів LLM-оцінювання з експертною оцінкою.

6. Орієнтовний перелік графічного (ілюстративного) матеріалу: 3 плакати формату А1
7. Орієнтовний перелік публікацій: не представлено
8. Консультанти розділів дисертації

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Розробка стартап-проектів	Бояринова К.О., д.е.н., проф. завідувач кафедри економічної кібернетики КПІ ім. Ігоря Сікорського	05.11.2025	

9. Дата видачі завдання 01.09.2025

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1	Формулювання завдання магістерської дисертації	01.09.2025	Виконано
2	Первинний аналітичний огляд теоретичного матеріалу (збір джерел, аналіз стану проблеми)	11.09.2025	Виконано
3	Ознайомлення з вимогами до оформлення та структурою магістерської дисертації	26.09.2025	Виконано
4	Формування детального плану роботи та уточнення дослідницьких задач	29.09.2025	Виконано
5	Підготовка експериментального середовища	06.10.2025	Виконано
6	Проведення експерименту та аналіз отриманих даних	29.10.2025	Виконано
7	Впровадження запропонованих покращень до робочого прототипу	02.11.2025	Виконано
8	Розробка стартап-проєкту (маркетинг, стратегія, бізнес-модель)	05.11.2025	Виконано
9	Підготовка остаточної документації та формування висновків	10.12.2025	Виконано

Студент М.М. Журавель

Науковий керівник А.Г. Протасов

Реферат

Актуальність теми: основним каналом комунікації клієнтів з бізнесом є контакт центр, а підтримка стандартів якості компанії і загалом якість спілкування є одним із ключових факторів у досягненні лояльності і репутації компаній, що приносить як задоволеність клієнту так і ефективність сервісу. Базовий контроль розмов вручну не може виконати 100% перевірку, а покриває лише невелику частину діалогів - як правило 10-15%, більше того часто така оцінка є суб'єктивною і залежить від навичок контролери і частоти командної синхронізації. Використання мовних моделей може не тільки автоматизувати аналіз, а й підвищити об'єктивність оцінки та масштабувати процес без необхідності додаткового персоналу. Все це робить дослідження актуальним і практичним.

Мета дослідження: удосконалення процесу контролю якості комунікацій у контакт-центрах шляхом розроблення прототипу автоматизованої системи оцінювання, що ґрунтується на багаторівневій моделі аналізу діалогів та застосуванні мовних моделей (LLM).

Завдання дослідження:

1. провести аналіз сучасних методів контролю за якістю розмов;
2. побудувати концептуальну модель системи контролю якості;
3. розробити багаторівневу модель оцінювання комунікацій, що включає критерії, підкритерії та формалізовані індикатори порушень;
4. обґрунтувати вибір алгоритмів і архітектурних рішень для автоматизованого аналізу діалогів;
5. розробити, реалізувати та експериментально перевірити MVP автоматизованої системи контролю якості шляхом порівняння результатів LLM-оцінювання з експертною оцінкою;
6. визначити напрями подальшого вдосконалення та масштабування системи.

Об'єкт дослідження: процес контролю якості комунікацій у контакт-центрах, що включає інформаційні, програмні та організаційні компоненти автоматизованих систем оцінювання.

Предмет дослідження: методи та алгоритмічні принципи автоматизованого аналізу якості комунікацій у контакт-центрах на основі багаторівневої моделі оцінювання із застосуванням мовних моделей (LLM) для виявлення й інтерпретації порушень у діалогах.

Наукова новизна дослідження:

- удосконалення методу аналізу розмов із формуванням доказової бази;
- застосування механізму доступу до бази знань і інструкцій для мовної моделі;
- обґрунтування методики перевірки точності на основі порівняння з еталонними результатами і наданими інструкціями по параметрам;

Практична цінність роботи:

- створено прототип автоматизованої системи контролю якості;
- реалізовано визначення порушень та пояснень;
- архітектура придатна для інтеграції у корпоративні системи та масштабування під різні канали комунікації;
- запропонування багаторівневої моделі для оцінки розмов.

Ключові слова: контакт-центр; контроль якості комунікацій; автоматизована система; багаторівнева модель оцінювання; підкритерійний аналіз; мовні моделі; аналіз діалогів; пояснюваність результатів; коучинговий зворотний зв’язок.

Abstract

Relevance of the study: contact centers are one of the primary channels of communication between customers and businesses, while maintaining service quality standards and overall communication quality is a key factor in achieving customer loyalty and corporate reputation, directly affecting customer satisfaction and service efficiency. Manual quality control of conversations is unable to provide full coverage, typically reviewing only 10–15% of dialogues. Moreover, such evaluations are often subjective and depend on the individual skills of quality controllers and the frequency of team calibration. The use of large language models enables not only the automation of dialogue analysis but also increases objectivity and scalability without the need for additional personnel. These factors determine the relevance and practical significance of the study.

Purpose of the study: to improve the process of communication quality control in contact centers by developing a prototype of an automated evaluation system based on a multi-level dialogue analysis model and the application of large language models (LLMs).

Research tasks:

1. to analyze modern approaches to communication quality control in contact centers;
2. to develop a conceptual model of a quality control system;
3. to design a multi-level communication evaluation model that includes criteria, subcriteria, and formalized indicators of violations;
4. to substantiate the selection of algorithms and architectural solutions for automated dialogue analysis;
5. to develop, implement, and experimentally validate an MVP of an automated quality control system by comparing LLM-based evaluation results with expert assessments;
6. to determine directions for further system improvement and scaling.

Object of the research: the process of communication quality control in contact centers, including the informational, software, and organizational components of automated evaluation systems.

Subject of the research: Methods and algorithmic principles of automated communication quality analysis in contact centers based on a multi-level evaluation model using large language models (LLMs) to detect and interpret violations in dialogues.

Scientific novelty of the research:

- improvement of dialogue analysis methods through the formation of an evidence-based evaluation framework;
- application of a knowledge base and instruction retrieval mechanism for large language models;
- substantiation of an evaluation accuracy verification methodology based on comparison with reference results and predefined instruction parameters;

Practical value of the work:

- a prototype of an automated communication quality control system has been developed;
- detection of violations and generation of explanatory feedback have been implemented;
- the proposed architecture is suitable for integration into corporate systems and scalable across different communication channels;
- proposal of a multi-level dialogue evaluation model.

Keywords: contact center; communication quality control; automated system; multi-level evaluation model; sub-criterion analysis; language models; dialogue analysis; explainability of results; coaching feedback.

ЗМІСТ

СПИСОК СКОРОЧЕНЬ І ТЕРМІНІВ

Технічні скорочення

АСКЯК	Автоматизована система контролю якості комунікацій.
АТС	Автоматична телефонна станція (PBX).
ШІ / AI	Штучний інтелект / Artificial Intelligence.
API	Application Programming Interface.
ASR	Automatic Speech Recognition.
CES	Customer Effort Score.
CoT	Chain-of-Thought.
CPU / GPU / RAM	Central Processing Unit / Graphics Processing Unit.
CRM	Customer Relationship Management.
CSAT	Customer Satisfaction.
CSV	Comma-Separated Values.
CX	Customer Experience.
GDPR	General Data Protection Regulation.
IRR	Inter-rater Reliability.
JSON	JavaScript Object Notation.
LLM	Large Language Model.
LoRA	Low-Rank Adaptation.
MVP	Minimum Viable Product.
NLP	Natural Language Processing.
NPS	Net Promoter Score.
QA / QCS	Quality Assurance / Quality Control of Service.
PEFT	Parameter-Efficient Fine-Tuning.
RAG	Retrieval-Augmented Generation.
RBAC	Role-Based Access Control.
RLHF	Reinforcement Learning from Human Feedback.
SLI / SLO	Service-Level Indicator / Service-Level Objective.

TLS / AES-256	Transport Layer Security / Advanced Encryption Standard 256-bit.
VIP	Very Important Person.
VPS / VDS	Virtual Private Server / Virtual Dedicated Server.

Аналітичні, процесні та бізнесові терміни

4E-модель (Effectiveness, Efficiency, Explainability, Ethicality)	інтегральна модель оцінювання ефективності системи: результативність, операційна ефективність, пояснюваність, етичність.
Байєсівські підходи (Bayesian methods)	статистичні методи оцінювання та оновлення ймовірностей; застосовуються для моделювання невизначеності та корекції метрик якості.
Датасети (datasets)	структуровані набори даних; використовуються для навчання, валідації та тестування моделей ШІ.
Квантизація (quantization)	оптимізація моделі шляхом зниження розрядності ваг для зменшення ресурсів, збереження прийнятної якості.
Омніканальність (omnichannel)	підтримка кількох каналів комунікації (voice, chat, email) з єдиною логікою аналізу.
Пайплайн (pipeline)	послідовність етапів обробки даних у системі (парсинг → аналіз → оцінювання → збереження результатів).
Парсинг (parsing)	технічний процес розбору структури даних (наприклад, CSV-файлу з діалогом).
Патерни (patterns)	повторювані поведінкові або лінгвістичні шаблони у діалогах, які вказують на типові помилки або сценарії.

Промпт (prompt)	інструкція або запит, що передається мовній моделі для керування логікою відповіді.
Скриншоти (screenshots)	графічні зображення інтерфейсу системи, що використовуються як ілюстрації в роботі.
Таймінги (timings)	часові характеристики діалогу: швидкість відповіді, паузи, затримки між репліками.
Фідбек (feedback)	зворотний зв'язок щодо оцінки якості розмови або результатів аналізу.
Чек-листи (checklists)	структуровані списки критеріїв чи вимог для аудиту розмов.
Accuracy (точність), Precision (прецизійність), Recall (повнота), F1-score	стандартні метрики якості класифікаційних моделей, що часто використовуються для валідації систем автоматичного скорингу.
Agent / Агент	оператор контакт-центру, який взаємодіє з клієнтами.
API-based / Self-hosted / Hybrid	варіанти розгортання системи: через зовнішній API, локально або у комбінованій архітектурі.
Block-level аналіз	аналіз діалогу на рівні великих блоків (комплаєнс, структура, емпатія), без деталізації на підкритерії.
Bolt-on інтеграція (bolt-on AI)	поверхнева інтеграція AI-модулів у вже існуючі системи без глибокої зміни архітектури.
Coaching feedback / Коучинговий фідбек	аналітичне резюме або рекомендації для агента за підсумками оцінки розмови.
Continuous QA	Continuous Quality Assurance, безперервний процес контролю якості з автоматичним зворотним зв'язком і самооновленням системи.
Correlation	показник узгодженості між автоматичними оцінками системи та людськими експертними оцінками.

Customer Experience Analytics	аналітика клієнтського досвіду, яка поєднує дані QCS, CRM та поведінкові метрики.
Data drift	зміна статистичних властивостей даних у часі, що впливає на стабільність роботи моделі.
Development / Staging / Production	Етапи середовища розробки: розроблення, тестування та експлуатація.
Distillation	перенесення знань від більшої (teacher) моделі до компактнішої (student) для покращення продуктивності.
Docker / Kubernetes	Платформа контейнеризації для ізольованого розгортання застосунків та відтворюваного середовища розробки.
Embeddings	Векторні представлення слів або речень, що використовуються LLM для обчислення семантичної подібності.
Endpoint	мережева адреса для виклику API-моделі або сервісу.
Error-level	рівень критичності помилки (minor/major) у системі оцінювання.
Evaluation plan	план/ієрархія підкритеріїв для відповіді на конкретне QA-питання; використовується для підвищення стабільності оцінки між моделями.
Evidence	доказові фрагменти діалогу / цитати.
Explainability (XAI)	Пояснюваність рішень системи, здатність моделі аргументувати отримані оцінки.
F1-score	метрика збалансованості між Precision і Recall; використовується для валідації моделей класифікації.
Fallback-механізми (fallback mechanisms)	резервні правила або моделі на випадок помилки основного методу аналізу.

Few-shot	режим роботи LLM, коли модель отримує кілька прикладів для формування логіки відповіді.
Fine-tuning / Донавчання	донавчання моделі на специфічних даних (наприклад, дзвінках певної компанії) для покращення релевантності.
GenAI	Generative AI; тип ШІ, здатний генерувати текст або інші дані на основі діалогів.
Global instruction prompts	універсальні глобальні інструкції для LLM, що задають базову логіку аналізу.
GPT-4 mini	компактна мовна модель GPT-4 класу, оптимізована для швидкого та бюджетного аналізу діалогів.
Human-in-the-Loop (HITL)	Людина в циклі, принцип включення експерта до процесу перевірки й навчання моделі.
Issue	перелік виявлених помилок або порушень за критеріями.
Latency	затримка у відповіді системи; важлива метрика в реальному використанні чат-моделей.
LLM-readiness	рівень готовності компанії або продукту до впровадження рішень на основі LLM.
LLM-підхід	використання LLM як ядра автоматизованої системи оцінювання.
Low-Rank Adaptation	метод компактного донавчання LLM на специфічних даних.
Machine learning / deep learning	методи машинного та глибинного навчання, що передували LLM-підходу.
Minor / Major	несуттєве / суттєве порушення
MySQL	реляційна база даних, у якій зберігаються діалоги та результати аналізу в MVP.

Prompt engineering	методика розробки, структурування й оптимізації промптів для стабільних і контрольованих відповідей.
Prompt optimization	підбір структури та формату промпта для підвищення точності моделі.
Reasoning	здатність моделі виконувати логічні міркування, а не просто генерувати текст.
Retention Rate	частка клієнтів, що повертаються; використовується як бізнес-метрика ефективності сервісу.
Retrieval-методи (retrieval methods)	методи вибірки релевантної інформації з бази знань у рамках RAG.
Rule-based	системи, які працюють на жорстких правилах і чек-листах без контекстного аналізу.
Scorecard / Скоркард	матриця оцінювання якості розмови за низкою критеріїв (комплаєнс, емпатія, структура, результат). Використовується для аудиту агентів.
Scoring / Скоринг	процес виставлення оцінок за розмову на основі скоркарду або моделі.
Sentiment/emotion detection	визначення тональності та емоційної складової у зверненнях клієнтів.

1. АНАЛІТИЧНИЙ ОГЛЯД

Метою цього розділу є проведення системного аналізу сучасного стану та тенденцій розвитку методів контролю якості комунікацій у контакт-центрах. Аналітичний огляд охоплює теоретичні засади, історичні передумови, наукові підходи та практичні рішення, що сформували сучасне уявлення про процес оцінювання ефективності роботи живих агентів. У сучасній практиці контакт-центрів для позначення процесів перевірки комунікацій між клієнтами та агентами часто використовується термін Quality Assurance (QA). Проте з позиції теорії управління якістю більш коректним у даному контексті є поняття Quality Control of Service (QCS) — тобто контроль якості обслуговування [27][31], що передбачає безпосередній аналіз розмов, виявлення відхилень від стандартів і формування рекомендаці. У цій роботі терміни QA та QCS вживаються як синоніми, з пріоритетом змістового значення — контроль якості комунікацій у контакт-центрі. Проведений аналіз дозволяє виявити закономірності розвитку галузі, ідентифікувати обмеження традиційних методів та обґрунтувати необхідність удосконалення технологічних рішень.

Тематика огляду зосереджена на вивченні еволюції систем перевірки якості обслуговування клієнтів — від ручних і rule-based методик до автоматизованих інтелектуальних систем, що базуються на технологіях розпізнавання мовлення, обробки природної мови (NLP) та великих мовних моделей (LLM). Увага приділяється питанням узгодженості між оцінками людини та машинними алгоритмами, об’єктивності аналітики, масштабованості систем і зв’язку між якістю комунікації та рівнем задоволеності клієнтів[25][26][29].

Результати цього розділу створюють теоретико-аналітичну основу для подальшого формулювання наукових завдань роботи. Вони дозволяють визначити, які методи й технологічні підходи є найбільш ефективними для розбудови автоматизованої системи контролю якості розмов живих агентів, що є предметом дослідження цієї магістерської роботи.

1.1.Загальні відомості про перевірку якості розмов живих агентів

Контроль якості обслуговування клієнтів у контакт-центрах є складовою системи управління сервісною взаємодією і клієнтським досвідом[28][29][30]. У міру зростання вимог до швидкості, точності та персоналізації комунікацій системи оцінювання роботи агентів стають не просто інструментом контролю, а функцією стратегічного управління сервісною взаємодією.

Традиційно оцінювання якості виконувалося вручну: аудитори вибірково прослуховували записи дзвінків або читали розмови в чатах/листах, фіксували відхилення від стандартів і формували рекомендації. Модель добре працювала в умовах обмежених обсягів комунікацій, але в епоху цифровізації, появи чатів та омніканальних звернень вона втратила масштабованість і об'єктивність. У практиці контакт-центрів ручний аудит розмов має вибірковий характер і зазвичай охоплює лише невелику частку діалогів (як правило, однозначні або низькі двозначні відсотки), що не дозволяє сформувати повну картину якості обслуговування [23][32][33][34].

Паралельно до цього змінився зміст самого поняття якості. Сьогодні контроль якості розглядається не тільки як відповідність скриптам чи регламентам, а як інструмент управління клієнтським досвідом (CX) [26][29]. Як зазначено в [3], аудитор аналізує не просто зміст відповідей, а логіку взаємодії, емоційний фон, дотримання стандартів, ефективність вирішення та інтонаційний контекст.

Перші системи якості (1990–2000) базувалися на чек-листах: аудитор оцінював структуру привітання, коректність інформації, логіку відповідей [31]. Наступний етап (2000–2010) пов'язаний з появою цифрових АТС (інфраструктура маршрутизації дзвінків у контакт-центрі) і перших інструментів speech analytics, здатних автоматично фіксувати ключові слова чи паузи. З 2010-х розвиток NLP і класифікаційних моделей дав змогу аналізувати наміри клієнтів, тематику розмови та типові патерни поведінки.

Суттєвий етап розвитку систем контролю якості пов’язаний з появою LLM (з 2023 року), що ознаменувало перехід до нового рівня складності систем контролю якості: моделі навчилися розуміти діалог у контексті, пояснювати свої рішення, працювати в zero-shot/few-shot режимах [37][38][39] та узгоджуватись з людськими оцінками на рівні до 0.6–0.7 за відомостями Spearman [1].

Описані етапи розвитку систем контролю якості відображають не лише хронологічні зміни, а й поступове зростання інтелектуальної складності підходів до аналізу комунікацій. На рисунку 1.1 наведено узагальнену схему еволюцію складності систем контролю якості — від rule-based рішень із мінімальним рівнем інтерпретації до сучасних LLM-based систем, здатних виконувати контекстний аналіз, логічне міркування та формувати результати.

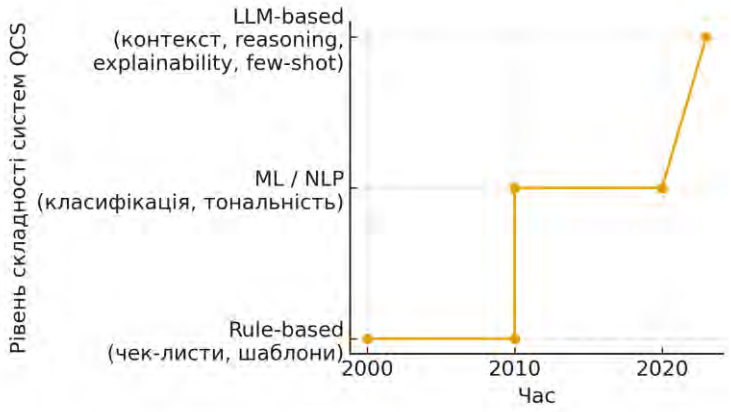

Рисунок 1.1. Графік еволюції складності систем контролю якості

Для наочного зіставлення основних поколінь систем контролю якості, їхніх функціональних можливостей та обмежень [24] у таблиці 1.1 наведено узагальнену класифікацію еволюції відповідних підходів.

Таблиця 1.1. Еволюція систем контролю якості

Покоління	Характеристика	Основні обмеження
Rule-based (2000–2010)	Чек-листи, словники фраз, жорсткі правила	Немає контексту, низька гнучкість
ML/NLP (2010–2020)	Класифікація тексту, тональність, тематика	Потребують великих датасетів, слабка explainability
LLM-based (2023+)	Контекстний аналіз, reasoning, пояснення, few-shot	Висока вартість, bolt-on інтеграції, нестабільність промптів

Проблеми традиційних підходів

Ручні та rule-based підходи мають низку системних обмежень:

- охоплення лише малої частки контактів → неповна картина якості;
- суб'єктивність оцінювачів → низька міжоцінювальна узгодженість [48];
- затримка фідбеку → неможливість оперативного коучингу;
- відсутність зв'язку з CRM-даними → немає персоналізації;
- неможливість масштабувати процес у великих контакт-центрах.

У звіті Call Criteria [2] підкреслюється, що відсутність стандартизованих підходів та «повільний» фідбек роблять класичні методи неефективними в умовах високої динаміки клієнтських звернень.

Еволюція вимог ринку та появи нових метрик

З розвитком CX-аналітики (інтегральне сприйняття сервісу клієнтом) акцент змістився:

- від фіксації порушень до вимірювання поведінкових патернів,
- від оцінки однієї розмови до впливу на бізнес-метрики,
- від універсальних стандартів до персоналізації залежно від сегменту клієнта.

Подібна трансформація підходів до контролю якості фіксується і в галузевих бенчмаркінгових оглядах Zendesk, де QA розглядається як інструмент управління клієнтським досвідом і бізнес-результатами, а не лише перевірки дотримання регламентів [23].

З огляду на зміну фокусу з формальної перевірки дотримання скриптів на комплексне оцінювання клієнтського досвіду, у сучасних системах контролю якості формується багатовимірний підхід до вимірювання ефективності комунікацій. Узагальнену структуру основних груп метрик, які використовуються для оцінювання якості діалогів у контакт-центрах, наведено в таблиці 1.2.

Таблиця 1.2. Групи метрик оцінювання якості діалогів

Група метрик	Опис
Операційні	ефективність процесу
Клієнтські	оцінка клієнтського досвіду
Комунікативні	поведінкові та структурні ознаки діалогу

Ця тришарова модель лягла в основу сучасних систем QCS і пояснює, чому автоматизація потребує складних інтелектуальних інструментів, здатних працювати з контекстом, емоційністю, поведінковими патернами, а не просто з текстом.

Вимоги до валідності та роль explainability

Сучасні системи мають відповідати вимогам:

- валідність — відповідність оцінки реальним критеріям якості;
- стабільність — мінімальна залежність від формулювань промптів;
- масштабованість — здатність покривати 100 % діалогів;
- пояснюваність — модель повинна аргументувати оцінки [47].

LLM уперше надали можливість автоматично генерувати пояснення до кожного висновку, що робить їх найбільш перспективним інструментом в автоматичному оцінюванні розмов.

Актуальність дослідження зумовлена поєднанням кількох ключових чинників, які формують сучасний контекст розвитку контакт-центрів:

1. Економічний чинник.

Компанії прагнуть оптимізувати витрати на аудит якості. Ручні перевірки охоплюють 1–2 % взаємодій, що робить процес дорогим і малоефективним. Автоматизація скорингу через LLM дає змогу масштабувати контроль до 100 % діалогів без пропорційного зростання витрат [42][44].

2. Організаційний чинник.

Контакт-центри працюють в умовах постійного збільшення обсягів комунікацій та переходу до омніканальності. У таких умовах традиційні підходи не здатні забезпечити ані швидкість зворотного зв’язку, ані потрібний рівень стабільності оцінок [43].

3. Технологічний чинник.

Еволюція від rule-based систем до моделей глибокого навчання та LLM відкрила можливість переходу від поверхневого до контекстного аналізу діалогів. Сучасні моделі [2][3] демонструють кореляцію з людськими оцінками та здатність відтворювати логіку експертного аналізу.

4. Соціальний чинник.

Клієнти очікують високого рівня сервісу, емпатії та персоналізації. Порушення цих очікувань безпосередньо впливає на показники CSAT (загальна задоволеність клієнтів), NPS (індекс лояльності клієнтів), утримання клієнтів та репутацію компанії. Контроль якості переходить від формальної процедури до частини системи управління клієнтським досвідом.

Таким чином, актуальність теми визначається необхідністю створення рішень, які одночасно забезпечують високу точність оцінювання, масштабованість, швидкість, пояснюваність та інтеграцію з корпоративними стандартами сервісу [45][46].

1.2. Тенденції автоматизації перевірки якості розмов у контакт-центрах

Зростання обсягів комунікацій у контакт-центрах, розвиток омніканальних платформ та очікування клієнтів щодо швидких і точних відповідей спричинили перехід від ручних перевірок до автоматизованих систем контролю якості[34][42][43]. Початкові інструменти автоматизації зосереджувалися переважно на технічних параметрах дзвінків – тривалості, паузах, темпі мовлення чи частоті звернень. Проте зі збільшенням складності клієнтських кейсів та різноманітності каналів комунікації виникла потреба у системах, здатних аналізувати контекст, зміст та поведінкові патерни агентів [23][42][44].

Цей перехід був зумовлений низкою технологічних, операційних і бізнесових чинників. Компанії вимагали інструментів, які могли б одночасно забезпечити масштабованість, точність, пояснюваність результатів і можливість інтеграції з CRMб що являє собою систему керування взаєминами з клієнтами і є джерелом атрибутів для персоналізації контролю та іншими операційними системами. На цьому етапі сформувалися основні напрями автоматизації, що визначили сучасний ландшафт систем QCS.

Стан розвитку технологій автоматизації

Першою хвилею автоматизації стали системи speech analytics, здатні автоматично транскрибувати дзвінки та аналізувати інтонацію, мовний темп, паузи і ключові слова. Пізніші рішення, основані на NLP та машинному навчанні, дозволили визначати наміри клієнтів, аналізувати тональність, виявляти порушення скриптів і класифікувати діалоги на “задовільні” та “незадовільні”. Однак ці системи потребували великих наборів розмічених даних, мали обмежену узагальнюваність та слабку пояснюваність рішень [49].

Наступний етап розвитку автоматизації пов’язаний із появою LLM, які здатні працювати в режимах zero-shot і few-shot, аналізувати складні взаємодії та формувати пояснювані висновки [37][38][39]. Згідно з дослідженням [3], застосування LLM у задачах контролю якості комунікацій у контакт-центрах забезпечує приріст до 18–19 % за показником Macro F1 порівняно з традиційними rule-based та класичними ML-підходами в аналогічних умовах оцінювання.

Ключові технологічні напрями автоматизації контролю якості

На сучасному етапі формуються кілька ключових технологічних напрямів, що утворюють основу інтелектуальних систем QCS:

- Speech analytics — автоматичне розпізнавання мовлення (ASR), виявлення емоційної тональності та просодичних характеристик.
- NLP-аналіз — класифікація змісту, семантичні моделі, тематичне моделювання, аналіз намірів.
- Machine learning / deep learning — бінарна і багатокласова оцінка розмов, визначення аномалій, сегментація звернень.
- LLM-підхід — контекстне оцінювання діалогу, reasoning, пояснення рішень (explainability), оцінювання підкритеріїв у few-shot режимі.
- Інтеграція з корпоративними даними — CRM-атрибути, історія клієнта, профіль контракту, що дозволяє персоналізувати оцінку [42][44].
- RAG-процеси — використання баз знань як когнітивного контексту, що дозволяє моделі зіставляти дії агента з еталонними сценаріями [39].

Ці напрями формують ядро технологій, яке дозволяє відмовитися від вибіркового аудиту та переходити до повного автоматичного охоплення 100 % діалогів. Для узагальнення та порівняльного аналізу основних технологічних підходів, що застосовуються в автоматизації контролю якості комунікацій, у таблиці 1.3 наведено їхню характеристику з точки зору функціонального призначення, переваг та обмежень у контексті QCS-систем.

Таблиця 1.3. Технологічні підходи в автоматизації QCS

Підхід	Суть	Переваги	Обмеження
ASR / speech analytics (розпізнавання мовлення й аналіз просодії)	розпізнавання мовлення, аналіз інтонацій	стабільність, швидкість	немає глибинного аналізу змісту
NLP / ML	класифікація тексту, тональність, тематичний аналіз	певна контекстність	залежність від великих датасетів
Deep learning	глибинні багатошарові моделі	стійкість і якість класифікації	низька пояснюваність (explainability)
LLM-based (контекстний аналіз діалогу)	контекстне логічне міркування (reasoning), пояснення, few-shot режим	найвища точність і узгодженість	вартість, нестабільність промптів
RAG	поєднання бази знань (KB) та генерації	висока фактичність	потреба у структурованій базі знань

Ринкові тенденції

Сучасний ринок рішень QCS переходить від часткового автоматичного аналізу до комплексних платформ, які одночасно виконують семантичний аналіз, емоційну діагностику, формування coaching feedback і інтеграцію з CX-показниками. Згідно з дослідженнями [2][3], компанії активно переходять до моделей, що:

- працюють із повним охопленням (100 %),
- забезпечують explainability,
- можуть працювати з багатьма каналами (дзвінки, чати, листи),
- використовують байєсівські чи LLM-підходи до reasoning,
- підтримують інтеграцію з CRM для адаптивного скорингу.

Крім того, поширюється тенденція до локального розгортання модулів — особливо квантованих моделей [5], які забезпечують баланс між якістю та економічністю.

Серед комерційних систем доцільно відзначити MaestroQA як приклад платформи, що поєднує автоматизований аналіз, навчання, контроль показників і формування коучингових рекомендацій. Наявність AI-модуля з кастомними промптами та інтеграціями (зокрема з CRM) ілюструє перехід від вибіркового аудиту до системного управління якістю. Це робить платформу показовим прикладом переходу від аудиту до інтелектуальної системи управління якістю.

У процесі розвитку систем контролю якості розмов можна виокремити три покоління технологічних підходів.

Перше покоління, так звані rule-based системи (2000–2010 рр.), ґрунтувалося на фіксованих правилах і словниках заборонених чи обов'язкових фраз. Такі рішення мали перевагу у прозорості, однак характеризувалися низькою гнучкістю та відсутністю здатності враховувати контекст розмови [4].

Друге покоління (2010–2020 рр.) спиралося на методи машинного навчання та обробки природної мови. На цьому етапі з'явилися системи, що використовували алгоритми класифікації, тематичне моделювання та аналіз тональності для визначення якості комунікацій. Хоча ці підходи забезпечили вищу точність і ширше охоплення, вони вимагали значних обсягів розмічених даних і складної підтримки моделей.

Починаючи з 2023 року, формується третє покоління рішень — LLM-based системи, які застосовують великі мовні моделі (Large Language Models) для контекстного аналізу діалогів. Їхньою перевагою є здатність інтерпретувати семантичні й прагматичні аспекти розмови, здійснювати оцінювання у режимі few-shot та zero-shot, а також пояснювати отримані результати природною мовою.

Послідовність переходу від rule-based підходів до ML/NLP-рішень і далі до LLM-based систем відображає зміну логіки автоматизації — від жорстких правил до контекстної інтерпретації діалогів. Як зазначається у роботі [50], , навіть за наявності автоматизованих моделей ключовою проблемою залишається коректність оцінювання діалогів, орієнтованих на виконання запиту клієнта, оскільки формальні показники не завжди відображають реальну ефективність його вирішення. Цю еволюцію технологічних підходів та їхні ключові етапи наочно представлено на рисунку 1.2, де показано кроки розвитку поколінь систем контролю якості.

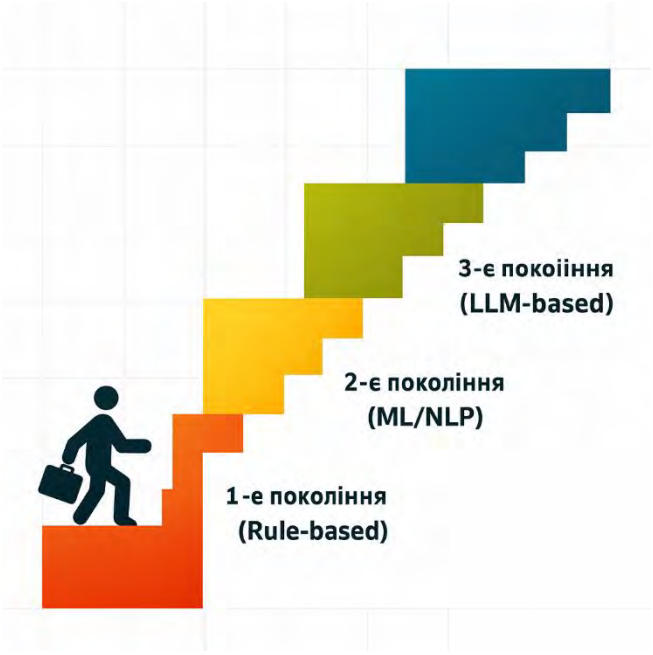

Рисунок 1.2. Технологічні підходи поколінь.

Перехід від rule-based перевірок до контекстного аналізу визначає вимоги до сучасних QCS-систем: інтерпретованість результатів, масштабованість і можливість інтеграції з корпоративними даними.

Сучасні платформи використовують комплексний підхід, що включає:

- послідовний конвеєр аналізу ASR → NLP → LLM reasoning як багаторівневий конвеєр аналізу,
- наліз емоційного стану та тональності висловлювань,
- онтекстне оцінювання якості діалогу, ,
- механізми пояснюваності отриманих результатів,

- підкритерійний аналіз (evaluation plans),
- доступ до корпоративних політик та KB (база знань — структурований набір даних для пошуку фактів і відповідей агентом) через RAG,
- коучингові пояснення у вигляді summary або рекомендацій.

Згідно з дослідженнями [1], найбільш валідними є системи, що поєднують reasoning + пояснення + покриття бази знань.

Серед поширених комерційних рішень, що реалізують автоматизацію QCS, можна виділити:

- KlausApp (Zendesk QA) — автоматичний скоринг, інтеграція зі Slack;
- Scorebuddy QA — Auto Scoring на основі GenAI, узгодженість до 95 % із людськими оцінками;
- Qualtrics CX — системний аналіз голосу клієнта;
- Evaluagent CX — гібрид Human + AI;
- Convin — LLM-аналіз поведінки агентів;
- Omind — емоційний аналіз і діагностика вигорання;
- MaestroQA — еталонна інтегрована система.

Попри швидкий розвиток, ці платформи все ще мають низку системних обмежень, які розглянуто в наступному підрозділі.

Тенденції подальшої еволюції систем QCS

Сьогодні розвиток рухається у напрямі:

- переходу від текстоцентричності до мультимодальності (зображення, файли, документи);
- інтеграції QCS та Customer Experience Analytics;
- переходу від «оцінки» до «рекомендацій і навчання»;
- впровадження Human-in-the-Loop для калібрування [51];
- зростання важливості explainability та валідації рішень;
- зменшення чутливості системи до змін формулювання інструкцій.

Таким чином, системи контролю якості переходять від інструментів перевірки до повноцінних інтелектуальних систем управління сервісною взаємодією.

1.3.Проблеми та обмеження сучасних систем автоматизованого контролю якості

Активне поширення LLM-рішень у сфері контролю якості комунікацій (Quality Control of Service) створило враження швидкого технологічного прориву. Проте більшість комерційних платформ реалізували інтеграцію за принципом bolt-on AI — тобто додали мовну модель поверх уже існуючої логіки ручного аудиту або простих rule-based механізмів. Такий підхід дозволив швидко відповісти на ринковий попит, але водночас виявив низку глибинних системних обмежень, які сьогодні визначають реальні межі автоматизації. Як показують результати досліджень [3], поверхнева інтеграція LLM не дає стабільної відповідності корпоративним стандартам, а результати моделі залишаються слабо відтворюваними у складних сценаріях спілкування.

Першою фундаментальною проблемою є те, що більшість платформ використовують LLM лише як модуль узагальнення: модель оцінює діалог глобально, але не співвідносить кожну репліку із конкретним підкритерієм корпоративного чек-листа. Замість структурного аналізу виникає узагальнена текстова «думка моделі», що унеможливлює формування відтворюваних результатів. Дослідження [1] підтверджує, що навіть найсильніші LLM демонструють нестабільність при зміні формулювання промпта та погіршення результатів у завданнях підкритерійного скорингу. Це і є причина, чому більшість рішень забезпечують лише block-level аналіз (комплаєнс, емпатія, структура), але не здатні розкласти порушення на атомарні дії агента.

Другим суттєвим обмеженням є залежність систем від одного універсального промпта. Використання “global instruction prompts” звучить привабливо з точки зору масштабованості, але різні типи звернень — технічна підтримка, продажі, білінг, претензії — потребують різної логіки аналізу. За

наявними даними [2], універсальні промпти призводять до втрати точності до 25–35 % у складних кейсах та роблять результати надмірно чутливими до контексту діалогу. У промислових умовах це спричиняє prompt drift: модель змінює стиль і логіку рішень у часі, що унеможливлює використання таких результатів для офіційної оцінки агентів.

Ще одним значущим обмеженням є слабка зв'язність із CRM-даними. У реальних контакт-центрах рівень сервісу має бути персоналізованим: VIP-клієнт та користувач із низьким тарифом мають різні очікування й різну цінність для бізнесу. Проте сучасні комерційні системи оцінюють якість без урахування атрибутів клієнта. Це призводить до однакової оцінки розмов із кардинально різними вимогами до сервісу. Підхід LLM без CRM-контексту створює ситуацію, коли оцінювання стає відірваним від бізнес-реальності. У дослідженнях [3] показано, що інтеграція клієнтських атрибутів підвищує точність скорингу, але більшість платформ досі не підтримують таку можливість.

Не менш критичною є технічна сторона. Через високу обчислювальну вартість моделі більшість систем змушені обмежувати довжину контексту. У результаті аналіз відбувається не на рівні повного діалогу, а на рівні коротких сегментів. Це порушує нитку взаємодії, руйнує причинно-наслідкові зв'язки та спотворює якість оцінювання. Проблему підтверджують деякі дослідження [5] при переході на локально розгорнуті квантовані моделі якість тримається лише тоді, коли система має доступ до повного RAG-контексту, а не до фрагментів.

Окрема група проблем пов'язана з мультимодальністю. У реальних діалогах клієнти часто надсилають фото, документи, файли чи скриншоти. Більшість QCS-платформ працюють виключно з текстом, відтинаючи значну частину інформації, необхідної для перевірки релевантності рішень агента. У результаті оцінювання втрачає валідність — система аналізує діалог без урахування ключових доказів чи вхідних даних. Це робить автоматизацію методологічно неповною.

Нарешті, ще одна проблема — відсутність навчальної придатності. Бізнес очікує, що QA-система підкаже агенту, що саме він зробив неправильно і як саме це виправити. Але більшість моделей повертають розмиті висновки («бракує емпатії», «некоректна ідентифікація потреб») без прив'язки до конкретних реплік. Відсутність error-level аналітики та timestamp-вказівок унеможливлює адресний коучинг і замкнутий цикл «оцінка → навчання → повторне вимірювання». Саме тому LLM-платформи досі не можуть повністю замінити людину-аудитора в частині розвиткового супроводу.

Розглянуті обмеження сучасних систем автоматизованого контролю якості мають системний характер і стосуються як архітектурних рішень, так і методології використання мовних моделей у промислових умовах. Для узагальнення виявлених проблем та їхнього впливу на результати оцінювання якості діалогів у таблиці 1.4 наведено основні обмеження сучасних LLM-систем контролю якості із зазначенням їхніх наслідків для практичного застосування.

Таблиця 1.4. Основні обмеження сучасних LLM-систем контролю якості

Проблема	Суть	Наслідки
Bolt-on інтеграція	LLM додано поверх старої логіки	поверхневе оцінювання, низька відтворюваність
Універсальний промпт	один шаблон для всіх типів звернень	падіння точності, prompt drift (дрейф промпта)
Відсутність CRM-контексту	немає персоналізації оцінки	некоректні висновки щодо «якості сервісу»
Обмеження контексту	скорочення діалогу через токени	втрата зв'язності аналізу
Немультимодальність	немає роботи з фото/документами	неповний аналіз складних кейсів
Відсутність error-level аналізу	немає точних вказівок на помилки	неможливість коучингу

Сумарно це формує ключовий висновок: попри швидке зростання можливостей LLM, більшість комерційних QCS-платформ залишаються лише частково автоматизованими системами. Вони покращують швидкість і охоплення, але не забезпечують ані еталонної точності, ані глибинної структурної аналітики. Саме це створює об'єктивну потребу у нових

архітектурних рішеннях, де LLM не є надбудовою, а виступає ядром контекстного, персоналізованого та керованого процесу контролю якості.

1.4. Наукова та практична актуальність дослідження

1.4.1. Наукова актуальність дослідження

Наукова актуальність даного дослідження зумовлена відсутністю формалізованих та відтворюваних методів автоматизованого контролю якості комунікацій між живими агентами у контакт-центрах із використанням великих мовних моделей (LLM). Незважаючи на активний розвиток досліджень у сфері діалогових систем, більшість наукових робіт зосереджена на оцінюванні взаємодії між людиною та штучним агентом, тоді як аналіз діалогів типу «людина — людина» у сервісному контексті залишається недостатньо дослідженим. За наявними публікаціями [1][3], застосування LLM для автоматизованого контролю якості людських діалогів у контакт-центрах перебуває на початковому етапі та не має усталеної методологічної бази.

У межах даної магістерської роботи запропоновано системний підхід до поєднання аналітичних, когнітивних і поведінкових критеріїв оцінювання якості обслуговування з алгоритмічними можливостями LLM. На відміну від існуючих підходів, орієнтованих переважно на поверхневу класифікацію або узагальнену оцінку діалогу, у роботі акцент зроблено на формалізації процесу аналізу комунікацій із урахуванням контексту, емоційних характеристик та логіки взаємодії учасників діалогу. Такий підхід відповідає сучасним напрямам досліджень у сфері людиноцентричного штучного інтелекту та дозволяє розширити можливості автоматизованого аналізу сервісних комунікацій.

Наукова новизна дослідження полягає у створенні передумов для:

- розроблення методології глибинного аналізу діалогів із використанням LLM, за якої мовна модель не лише ідентифікує порушення, а й відтворює логіку міркування експерта-аудитора;

- побудови ієрархічної системи критеріїв якості обслуговування — від окремих поведінкових індикаторів до узагальнених корпоративних стандартів, що підвищує валідність та інтерпретованість автоматичних оцінок;
- подальшого розвитку теорії автоматизованих систем підтримки рішень у сфері Quality Control of Service (QCS), які враховують контекст діалогу, емоційні характеристики та інтенції його учасників.

Отже, наукова цінність даної роботи полягає у формалізації підходу до інтеграції великих мовних моделей у системи оцінювання якості комунікацій, що створює методологічну основу для підвищення відтворюваності, валідності та практичної застосовності автоматизованого контролю якості в контакт-центрах.

1.4.2. Практична значущість дослідження

Практична значущість даного дослідження визначається потребами сучасних контакт-центрів у підвищенні ефективності процесів контролю якості обслуговування та зниженні витрат на ручний аудит. У більшості реальних умов частка перевірених вручну діалогів не перевищує 1–2 %, що не дозволяє отримати репрезентативну картину якості сервісу та своєчасно виявляти системні порушення у роботі агентів. Застосування автоматизованих систем на основі великих мовних моделей (LLM) створює можливість масштабування контролю якості до повного охоплення комунікацій, скорочення часу формування зворотного зв’язку та підвищення об’єктивності оцінювання.

Досвід упровадження систем, подібних до MaestroQA та Evaluagent, свідчить, що такі рішення суттєво зменшують навантаження на контролерів, але водночас залишають невирішеними проблеми глибини аналізу та персоналізації результатів. Саме ці недоліки роблять актуальним створення нового підходу, який передбачає:

- інтеграцію ШІ з корпоративними стандартами якості, тобто не просто застосування загальних промптів, а навчання моделі на базі реальних чек-листів компанії;
- зв’язок аналітики з клієнтськими атрибутами (CRM-даними), що дозволяє відрізняти обслуговування звичайних і VIP-клієнтів;
- поглиблену діагностику компетенцій агентів, коли LLM не лише констатує помилку, а пояснює її причину та вплив на бізнес-метрики (CSAT, NPS, Retention Rate);
- розширення можливостей коучингу — автоматичне формування індивідуальних рекомендацій на основі результатів QA-аналізу.

Практична користь дослідження проявляється у створенні передумов для розроблення MVP-системи (мінімально життєздатний прототип системи, що реалізує базові функції для перевірки концепції), описаної у завданні магістерської роботи, що дозволить інтегрувати інтелектуальний модуль контролю якості в реальні бізнес-процеси контакт-центру. Очікуваний результат — зниження витрат на ручний аудит, підвищення швидкості ухвалення рішень і формування єдиних стандартів оцінки для різних команд і каналів комунікації.

Висновки до розділу 1

Проведений аналітичний огляд показав послідовну еволюцію підходів до контролю якості комунікацій у контакт-центрах — від ручного аудиту та rule-based систем до сучасних LLM-орієнтованих рішень, здатних здійснювати контекстний, пояснюваний та масштабований аналіз діалогів. Попри значні технологічні зрушення, актуальні платформи залишаються обмеженими через поверхневу інтеграцію мовних моделей, відсутність персоналізації, неповну роботу з мультимодальністю й недостатню відтворюваність оцінок. Це формує об’єктивну потребу у створенні нових архітектурних підходів, які поєднують можливості LLM, RAG-механізмів та корпоративних стандартів якості.

2. ТЕОРЕТИЧНІ ОСНОВИ ПОБУДОВИ АВТОМАТИЗОВАНОЇ СИСТЕМИ КОНТРОЛЮ ЯКОСТІ КОМУНІКАЦІЙ

2.1. Формулювання наукової проблеми та постановка задачі

Розвиток систем контролю якості комунікацій у контакт-центрах пройшов еволюцію від вибіркових ручних перевірок до інтелектуальних рішень, заснованих на великих мовних моделях (LLM). Попри автоматизацію процесів, сучасні системи залишаються обмеженими: інтеграція LLM здебільшого поверхнева, результати характеризуються низьким рівнем пояснюваності, а персоналізація й мультимодальність недостатні. Це формує наукову проблему: як побудувати автоматизовану систему контролю якості комунікацій, що забезпечує достовірну, пояснювану та персоналізовану оцінку розмов за збереження економічної ефективності та масштабованості.

Теоретично така система розглядається як кібернетичний контур зворотного зв'язку, який виконує безперервне вимірювання, аналіз, корекцію й навчання. Вона повинна обробляти великі обсяги даних, оцінювати розмови за корпоративними стандартами, формувати числові показники й текстові пояснення, інтегруючи аналітику у процес коучингу агентів.

Для реалізації цього підходу необхідно поєднати сучасні інструменти штучного інтелекту (ШІ) — NLP (обробка природної мови та методи аналізу змісту транскриптів), LLM, контекстне навчання та RAG (підхід, що поєднує генерацію з вибіркою знань із бази для підвищення фактичної коректності), які формують базові функціональні властивості інтелектуальної системи контролю якості. Сукупність зазначених інструментів дозволяє сформувати систему, здатну забезпечувати контекстний аналіз діалогів, адаптацію до змін стандартів, пояснюваність результатів та відтворюваність оцінювання. Узагальнену модель поєднання сучасних інструментів ШІ, що відповідає сформульованій науковій проблемі, наведено на рисунку 2.1.

Рисунок 2.1. Поєднання сучасних інструментів ШІ.

Основні принципи побудови системи, що випливають з концептуальної моделі (рисунок 2.1):

- Контекстність — аналіз повного змісту розмови, а не окремих фраз;
- Пояснюваність (Explainability) — формування текстових аргументів до кожної оцінки;
- Адаптивність — урахування атрибутів клієнта й типу звернення;
- Відтворюваність — стабільність результатів за змінних умов;
- Економічність — використання оптимізованих (квантованих або локальних) моделей.

Практична реалізація цих принципів потребує балансу між аналітичною точністю великих моделей і доступністю для бізнесу з обмеженими ресурсами. Як показано у дослідженнях [5], поєднання квантованих моделей із RAG-архітектурою дозволяє досягати якості, співставної з API-рішеннями типу GPT-4 mini, за суттєво нижчої вартості. Тому у цій роботі розробляється мінімально життєздатний прототип (MVP), що реалізує базовий контур системи — від отримання даних до автоматичного оцінювання і зворотного зв’язку.

Розроблений MVP використовується для експериментальної перевірки працездатності ключових алгоритмів і підтвердження коректності запропонованого підходу. Його завдання:

- приймати вхідні транскрипти дзвінків або чати;

- здійснювати контекстний аналіз діалогу за допомогою LLM-модуля, інтегрованого у RAG-конвеєр;
- формувати автоматичну оцінку якості (scoring) за визначеними критеріями;
- генерувати коротке аналітичне пояснення (summary / feedback).

Результати порівнюються з еталонними зразками, що зберігаються у базі знань компанії або в стандартизованих промптах. У такий спосіб реалізується замкнений контур контролю якості.

У межах мінімально життєздатного прототипу цей контур реалізується у вигляді послідовного алгоритму автоматичного контролю якості, що охоплює етапи обробки вхідних даних, інтелектуального аналізу діалогу, звірки з еталонними критеріями та формування підсумкового звіту. Візуальне представлення цього алгоритму наведено на рисунку 2.2.

Рисунок 2.2. Алгоритм контролю якості розмов.

Результати реалізації MVP свідчать про можливість побудови автоматизованого контролю якості в межах мінімальної інфраструктури.

Виходячи з визначеної проблеми, завдання магістерської роботи полягає у розробленні теоретичних і методологічних засад побудови автоматизованої системи контролю якості комунікацій у контакт-центрах на основі LLM і RAG. Для досягнення мети необхідно:

- Сформувати концептуальну модель системи й визначити її ключові компоненти.
- Обґрунтувати вибір алгоритмів аналізу діалогів (LLM, RAG, квантування, CoT).

- Розробити архітектуру MVP, що забезпечує автоматичний скоринг і пояснення результатів.
- Визначити метрики ефективності (Accuracy, F1-score, Correlation, Explainability).
- Окреслити напрямки масштабування системи — інтеграцію з CRM, fine-tuning і мультимодальність.

2.2. Концептуальні засади побудови автоматизованої системи контролю якості комунікацій

Концепція автоматизованої системи контролю якості комунікацій базується на уявленні про процес оцінювання якості сервісної взаємодії як кібернетичний контур, що забезпечує безперервний цикл спостереження, аналізу, порівняння зі стандартами та формування зворотного зв'язку. На відміну від класичних моделей контролю, де ядром системи виступає людський аудитор, сучасні архітектури використовують інтелектуальні моделі, здатні виконувати семантичний аналіз діалогів, відтворювати логіку експерта та інтегрувати знання з корпоративних баз стандартів. Перехід до такої системної моделі зумовлений тим, що сучасні контакт-центри працюють у середовищі високої динаміки й масштабності, де ручні підходи не здатні забезпечити ні потрібний рівень охоплення, ні відтворюваність оцінок [2].

З позиції системного аналізу автоматизована система контролю якості комунікацій (АСКЯК), що являє собою програмно-аналітичну систему для оцінювання діалогів у контакт-центрі з використанням ШІ та формалізованих критеріїв, складається з декількох функціональних блоків, які взаємодіють у спільному управлінському контурі. Її основу становить модуль збору та підготовки даних, який приймає текстові чати, транскрипти дзвінків, метадані клієнтів і сервісних подій. Технології Automatic Speech Recognition забезпечують конвертацію аудіо в текст, тоді як додаткова нормалізація, очищення та сегментація готують матеріал до подальшої обробки. На цьому

етапі формується первинний корпус діалогів, який є вхідним шаром системи та визначає подальшу якість аналізу.

Аналітико-оцінювальний модуль системи виконує центральну функцію — семантичну інтерпретацію діалогу та формування оцінки відповідності поведінки агента корпоративним стандартам. На цьому етапі застосовуються великі мовні моделі (LLM), інтегровані з Retrieval-Augmented Generation (RAG), що дає змогу моделі зіставляти зміст діалогу з еталонними формулюваннями, політиками й сценаріями, збереженими в базі знань. Підхід RAG суттєво підвищує фактичну достовірність аналізу: замість того, щоб покладатися на внутрішні статистичні знання моделі, система здійснює фактичне звіряння тексту розмови з релевантними документами компанії. У проглянутих роботах [1][5] продемонстровано, що LLM у поєднанні з retrieval-методами демонструють значно вищу коректність і узгодженість з людською оцінкою, ніж моделі, які працюють у чисто генеративному режимі. Таким чином, ефективне автоматизоване оцінювання якості діалогу можливе лише за умови поєднання семантичних можливостей LLM, механізмів доступу до еталонних знань та структурованої бази корпоративних стандартів. Для наочного відображення взаємодії цих підходів та обмежень їх ізольованого застосування використано діаграму Венна, наведену на рисунку 2.3.

Рисунок 2.3. Діаграма Венна, як приклад поєднання методів.

У межах цього аналітичного блоку модель виконує кілька взаємопов’язаних функцій: семантичний аналіз намірів, оцінювання поведінкових патернів агента, визначення ключових порушень та формування первинного reasoning. На відміну від rule-based систем, де оцінювання має бінарний характер (є порушення — немає порушення), LLM здатні враховувати тональні нюанси, інтенції, емоційне забарвлення, логіку ведення діалогу та відповідність очікуванням клієнта — тобто здійснювати комплексну когнітивну оцінку [3]. Саме завдяки цьому LLM-системи, при коректній архітектурі, забезпечують кореляцію з людськими оцінювачами на рівні 0.6–0.7 за Spearman [1] і є найбільш перспективними інструментами для автоматичного QA.

Результати аналізу передаються в модуль пояснюваності (explainability), який формує текстове обґрунтування кожної оцінки, пояснюючи, які репліки діалогу стали причиною позитивної або негативної оцінки. Explainability виконує дві ключові функції: забезпечує прозорість системи й підвищує довіру до результатів, а також є основою для коучингової аналітики, оскільки дозволяє ідентифікувати конкретні дії агента та сформувати персоналізовані рекомендації. У роботах [6] зазначено, що поєднання chain-of-thought reasoning і пояснювальних відповідей суттєво покращує валідність LLM-рішень у задачах оцінювання.

Ключовим елементом концепції є база знань — когнітивне ядро системи, у якому містяться стандарти, політики, чек-листи, типові сценарії і приклади еталонної поведінки. У RAG-процесі база знань виступає джерелом еталонів, зіставлення з якими визначає якість конкретної взаємодії. Саме тому для побудови валідної АСКЯК вимоги до структурування та повноти бази знань є критичними. Дослідження [5] продемонструвало, що якість відповідей LLM у RAG-системах прямо залежить від "LLM-readiness" бази знань — чіткості формулювань, узгодженості термінології, відсутності дублювань та наявності прикладів.

Усі зазначені модулі утворюють єдиний інформаційно-аналітичний контур, у межах якого дані послідовно проходять етапи обробки, аналізу, оцінювання та формування зворотного зв'язку. Узагальнене подання цього контуру наведено на рисунку 2.4:

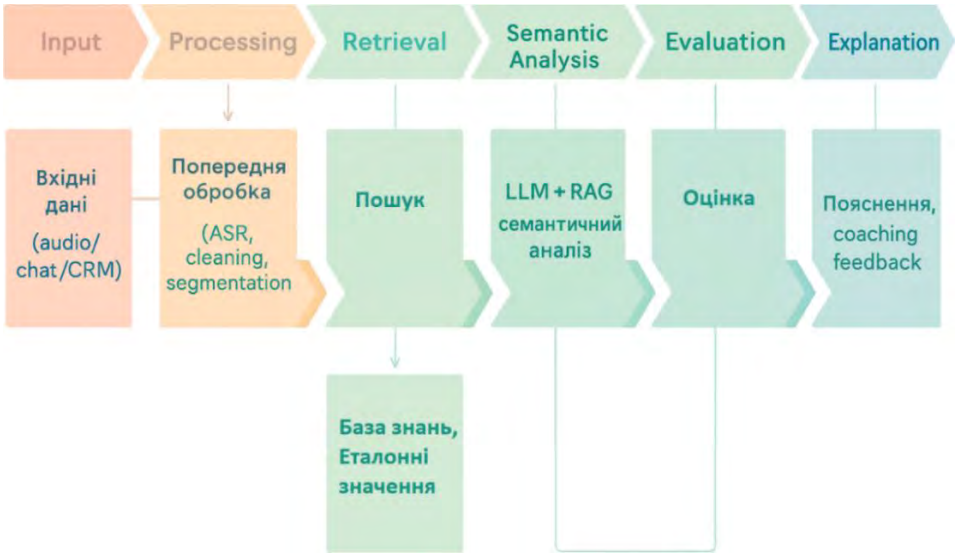

Рисунок 2.4. Узагальнений контур даних єдиної схеми контролю

Подана схема відображає узагальнений інформаційний контур роботи автоматизованої системи контролю якості комунікацій. На першому етапі система отримує вхідні дані — аудіозаписи, текстові чати або інформацію з CRM. Після цього відбувається попередня обробка, яка включає розпізнавання мовлення (ASR), очищення та нормалізацію тексту. Наступним кроком є пошук релевантних знань у базі знань компанії за допомогою векторних представлень, що забезпечує фактичну точність аналізу. Семантичний аналіз здійснюється великою мовною моделлю (LLM), доповненою механізмом Retrieval-Augmented Generation (RAG), що дозволяє зіставляти поведінку агента з корпоративними стандартами. Після цього система формує оцінку відповідності діалогу встановленим критеріям якості. Завершальним етапом є генерація пояснення та коучингових рекомендацій, що забезпечує прозорість оцінювання та підтримує процес професійного розвитку агентів.

У концептуальному аспекті система функціонує як багаторівнева модель управління якістю, де кожен цикл оцінювання доповнює корпоративну аналітику та створює умови для continuous QA — безперервного

вдосконалення стандартів, сценаріїв і дій агентів. Завдяки інтеграції CRM-атрибутів система здатна враховувати тип клієнта та контекст запиту, забезпечуючи персоналізоване оцінювання — той компонент, якого переважно бракує сучасним bolt-on платформам [3].

Узагальнюючи, концептуальні засади АСКЯК визначають її як інтелектуальний керований контур із п’яти ключових компонентів: збору даних, аналітичного reasoning, RAG-доступу до знань, пояснюваності та модулів навчання. Поєднання цих елементів утворює основу для побудови не просто автоматизованого скорингу, а системи експертного рівня, що здатна адаптуватися до корпоративних стандартів, забезпечувати відтворюваність результатів та інтегруватися в екосистему управління клієнтським досвідом.

2.3. Теоретичні основи роботи мовних моделей (LLM)

Сучасні автоматизовані системи контролю якості комунікацій у контакт-центрах базуються на здатності великих мовних моделей (Large Language Models, LLM) виконувати складні когнітивні завдання: розуміти контекст діалогу, розпізнавати інтенції, визначати релевантність висловлювань, виявляти порушення стандартів, робити висновки й пояснювати їх. На відміну від класичних rule-based або навіть традиційних NLP-систем, LLM працюють не з окремими фрагментами розмови, а з її структурою, логікою та поведінковими патернами, забезпечуючи рівень аналізу, максимально наближений до мислення експерта-аудитора [1]. Зазначені можливості великих мовних моделей не є абстрактними характеристиками, а можуть бути формалізовані у вигляді конкретних когнітивних властивостей, що безпосередньо впливають на якість автоматизованого аналізу діалогів у контакт-центрах. Саме ці властивості визначають здатність LLM виконувати контекстний, пояснюваний та відтворюваний контроль якості комунікацій у межах підходу Quality Control of Service (QCS). Узагальнення ключових когнітивних характеристик LLM та їх релевантності для задач контролю якості наведено у таблиці 2.1.

Таблиця 2.1. Ключові когнітивні властивості LLM, релевантні для QCS.

Властивість	Опис	Релевантність для QCS
Контекстність (global attention)	Аналіз зв’язків між усіма частинами діалогу, а не фрагментами	Дає змогу бачити логіку розмови, а не окремі порушення
Reasoning (ланцюги міркування)	Побудова логічної аргументації, CoT	Формування пояснення “чому агент порушив стандарт”
Емоційно-семантичне розуміння	Визначення намірів, тональності, емоцій	Аналіз емпатії, ввічливості, деескалації
Few-shot / zero-shot адаптація	Можливість працювати з мінімумом прикладів	Дає змогу налаштовувати оцінювання під конкретну компанію без великих датасетів
Генерація пояснень (explainability)	Формування природномовних обґрунтувань	Потрібно для коучингу та прозорості QA
RAG-сумісність	Поєднання генерації з пошуком у KB	Підвищує точність і знижує галюцинації

Теоретична основа LLM — це архітектура [7], у якій ключовим елементом є механізм багатоголовної уваги (multi-head attention). На відміну від рекурентних мереж, Transformer дозволяє моделі враховувати всі залежності в тексті одночасно, незалежно від дистанції між токенами. Завдяки цьому LLM формують глобальне семантичне представлення діалогу, що є критичним для задач аудиту якості: модель не просто «бачить» окремі порушення, а аналізує логіку розмови, її структуру, емоційний настрій та причинно-наслідкові зв’язки.

Процес роботи мовної моделі включає кілька етапів, які в сукупності забезпечують високий рівень когнітивного розуміння.

Спершу текст кодується у векторному просторі через embeddings — щільні числові представлення, які зберігають семантичні подібності між фразами. Далі Transformer-механізм визначає важливі зв’язки між частинами тексту, оцінюючи їхню вагу на основі контексту. Після цього модель виконує генерацію — побудову відповіді або інтерпретації, яка логічно узгоджується з усім діалогом.

Методи навчання LLM базуються на самопередбаченні: модель навчається прогнозувати наступний токен у тексті, опрацьовуючи сотні мільярдів слів, що забезпечує здатність до узагальнення в zero-shot і few-shot режимах [35]. Подальші стадії включають fine-tuning на специфічних корпусах та RLHF [6], який дозволяє адаптувати модель до людських уявлень про «правильні» відповіді. Саме RLHF надає LLM здатність до конструктивної аргументації, пояснень та відповідності корпоративному стилю.

У контексті контролю якості комунікацій важливе значення має здатність LLM виконувати zero-shot і few-shot навчання. Це означає, що модель може аналізувати новий тип діалогу або новий критерій оцінювання без попереднього глибокого навчання — достатньо правильно сформульованої інструкції або надання одного-двох еталонних прикладів. Ця властивість суттєво знижує бар'єр для впровадження систем, які використовують LLM, оскільки компанії не потрібно формувати великі корпуси розмічених діалогів — модель може працювати на основі скриптів, політик і стандартів.

Одним із найбільш ефективних методів підвищення валідності мовних моделей є Retrieval-Augmented Generation (RAG). Підхід RAG об'єднує генеративні можливості LLM із пошуком фактів у корпоративній базі знань, що дозволяє уникати галюцинацій, підвищувати точність і забезпечувати відповідність корпоративним політикам. У цьому разі модель отримує не лише текст розмови, а й додаткові релевантні документи, еталонні сценарії або чек-листи, що підключаються на етапі генерації оцінки. Як показано у Call Criteria [2], наявність RAG-доступу до політик суттєво підвищує фактичну коректність оцінювання, особливо у складних сценаріях з великою кількістю доменної інформації.

Важливою складовою є також оптимізація LLM для продуктивного та економічного використання в умовах обмежених ресурсів, що підтверджується сучасними підходами до побудови ефективних мовних моделей [36]. У практиці корпоративних систем QCS найпоширенішими є три напрямки оптимізації: квантизація, передавання знань від великої моделі до меншої

більш ефективної та PEFT. Квантизація (зниження розрядності ваг до 4–8 біт) дозволяє запускати моделі локально з високою швидкістю та низькою латентністю, не втрачаючи критичної точності [5]. Distillation дає змогу переносити знання від великих моделей до компактніших, що особливо важливо у системах з великою кількістю паралельних запитів. PEFT передбачає донавчання лише обмеженої підмножини параметрів моделі або додаткових адаптаційних шарів, що дозволяє адаптувати мовну модель до корпоративних чек-листів і стандартів без повного перенавчання та без зміни її базових параметрів. PEFT (наприклад, LoRA) дозволяє донавчати модель під конкретні корпоративні чек-листи, не змінюючи її основних параметрів.

У контексті задач контролю якості LLM демонструють здатність відтворювати логіку міркування експерта. Роботи [1] показують, що кореляція між LLM і експертними оцінками може перевищувати 0.8 для низки критеріїв. Застосування Chain-of-Thought підвищує пояснюваність і дозволяє моделі обґрунтовувати свої висновки у формі покрокових аргументів. Для контролю якості це особливо важливо: прозоре reasoning підвищує довіру аудиторів і слугує основою для коучингового аналізу.

Таким чином, мовні моделі формують теоретичну основу інтелектуальних систем контролю якості. Вони забезпечують контекстне, пояснюване та адаптивне оцінювання діалогів, здатне враховувати як лінгвістичні, так і поведінкові параметри взаємодії. У поєднанні з RAG-підходом і техніками оптимізації LLM стають центральним елементом системи контролю якості, яка здатна працювати з корпоративними стандартами, забезпечувати об’єктивність оцінювання та підтримувати безперервний цикл удосконалення сервісної комунікації.

2.4.Архітектурні принципи побудови системи контролю якості комунікацій

Архітектура автоматизованої системи контролю якості комунікацій визначає логіку взаємодії всіх модулів, що забезпечують обробку діалогів,

аналіз їх змісту, оцінювання відповідності корпоративним стандартам і формування пояснювального висновку. На відміну від традиційних платформ, які реалізують поверхневу bolt-on інтеграцію мовної моделі [3], сучасна архітектура повинна розглядатися як багаторівнева система, у якій LLM є центральним когнітивним ядром, а RAG забезпечує зв'язок із корпоративною базою знань.

2.4.1. Багаторівнева архітектура системи

Архітектура АСКЯК реалізована у вигляді чотирьох взаємопов'язаних рівнів:

1. Рівень даних (Data Layer).

Відповідає за збір і зберігання інформації про взаємодії — аудіо, чати, e-mail, CRM-звернення. Передбачає автоматичну транскрипцію (ASR), очищення, нормалізацію й підготовку текстів до аналізу.

2. Рівень обробки (Processing Layer).

Виконує базову структуризацію тексту, виділення тем і ключових елементів, побудову векторних представлень (embeddings) для пошуку релевантних знань у RAG-пайплайні.

3. Інтелектуальний рівень (AI Layer).

Ядро системи, де LLM, інтегрована з RAG, аналізує зміст діалогів, співставляє їх із корпоративними стандартами й формує скорингові оцінки з поясненнями.

Модель функціонує як експерт-аудитор, що ідентифікує сильні й проблемні аспекти розмови.

4. Рівень зворотного зв'язку (Feedback Layer).

Генерує короткі пояснення та рекомендації, передає результати до панелі контролю або системи коучингу. Забезпечує безперервне вдосконалення (Continuous QA).

Така архітектура модульна, масштабована й може інтегруватися з корпоративними платформами управління клієнтським досвідом (CEM).

2.4.2. MVP як ядро архітектури системи

У межах магістерського дослідження реалізується мінімально життєздатна версія системи (MVP), яка виконує основні функції архітектури, забезпечуючи замкнутий цикл «вхід — аналіз — оцінка — пояснення».

MVP складається з таких функціональних блоків:

- Data ingestion module — прийом і збереження транскриптів дзвінків або текстових чатів.
- Preprocessing module — очищення, токенізація й нормалізація тексту.
- LLM core — мовна модель, інтегрована через RAG, яка проводить семантичний аналіз, визначає релевантні знання з бази та здійснює скоринг.
- Evaluation module — порівняння отриманих оцінок із еталонними, заданими у промпті або базі знань.
- Explainability module — формування текстових пояснень результату.
- Output interface — формування підсумкового звіту з оцінкою, коротким самарі та рекомендаціями.

Загальна логіка роботи MVP як ядра архітектури автоматизованої системи контролю якості може бути представлена у вигляді послідовного контуру обробки даних — від надходження вхідної інформації до формування оцінки та пояснювального зворотного зв’язку. Такий підхід дозволяє наочно відобразити місце кожного функціонального модуля в загальному процесі аналізу та підкреслити роль еталонно-нормативного порівняння як центрального елемента системи. Архітектурну логіку MVP у спрощеному вигляді наведено на рисунку 2.4.

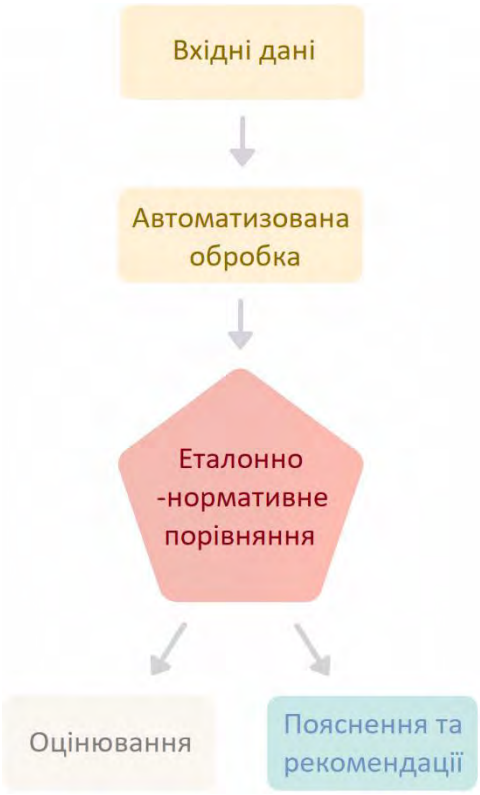

Рисунок 2.5. Архітектурна логіка MVP.

Як показано на рис. 2.5, архітектурна логіка MVP реалізує принцип прийняття вхідних даних, їх автоматизованої обробки, еталонно-нормативного порівняння та формування результату у вигляді оцінки якості й рекомендацій для агента.

На рівні MVP система працює з текстовими даними, без мультимодальності (аудіо, відео), проте підтримує повний цикл аналітики й зворотного зв’язку, що дозволяє підтвердити працездатність концепції.

Ключовим елементом архітектури є інтеграція RAG-підходу, що забезпечує динамічне звернення до бази знань під час кожної оцінки. Процес взаємодії виглядає таким чином:

1. Транскрипт або текст діалогу надходить до LLM-модуля.
2. Модель створює векторне представлення (embedding) ключових фрагментів тексту.
3. Вектор порівнюється з даними у базі знань для пошуку релевантних політик чи еталонів.

4. Знайдені фрагменти передаються моделі як контекст для формування висновку.
5. LLM здійснює оцінювання діалогу за критеріями.
6. Результат оцінки передається до evaluation module, де порівнюється з еталонною шкалою (інструкція або стандарт QA).
7. Модуль пояснення формує текстове обґрунтування та короткий summary.

Таким чином, система не обмежується статистичним аналізом, а використовує семантичне порівняння з нормативними знаннями, що забезпечує глибину й об’єктивність оцінки.

Для узагальнення функціонального наповнення MVP доцільно структурувати ключові модулі системи та їх призначення у вигляді таблиці 2.2.

Таблиця 2.2 — Функціональні модулі MVP-системи контролю якості.

№	Модуль системи	Основне призначення	Результат роботи
1	Модуль збору даних	Прийом текстових або транскрибованих діалогів з контакт-центру	Структурований діалог
2	Модуль автоматизованої обробки	Нормалізація, сегментація, підготовка діалогу до аналізу	Підготовлений текст
3	Інтелектуальний модуль (LLM)	Семантичний аналіз реплік та поведінкових патернів	Виявлені відхилення
4	Модуль перевірки відповідності	Порівняння з еталонами та корпоративними стандартами	Перелік порушень
5	Модуль оцінювання	Формування кількісної та якісної оцінки	Підсумкова оцінка
6	Модуль пояснень	Генерація обґрунтувань і рекомендацій	Пояснювальний висновок

2.4.3. Інфраструктура для реалізації MVP

Для розгортання MVP використовується платне хостинг-середовище з повним адміністративним контролем (VPS/VDS-рівень — віртуальний сервер для автономного хостингу MVP-системи), що забезпечує ізоляцію середовища виконання, контроль доступу до даних та стабільну роботу сервісів MVP.

Архітектурна конфігурація MVP відповідає стандартам серверних рішень для веб-додатків і включає такі компоненти: Операційне середовище, Веб-сервер, База даних, Інтерфейс управління.

Таке середовище забезпечує повну сумісність із бібліотеками для роботи з LLM-модулями через API, з що являють собою прикладний програмний інтерфейс для обміну даними між модулями або системами або через локальні інстанси моделей. Для зберігання знань, стандартів і промптів може використовуватись реляційна база даних MySQL, яка містить:

- таблиці корпоративних стандартів обслуговування (QA-checklists);
- приклади діалогів і скоринг-матриці;
- параметри оцінювання якості;
- шаблони промптів для мовної моделі.

Така конфігурація інфраструктури забезпечує відтворюваність експериментів, оперативне оновлення компонентів та можливість масштабування MVP до промислового рівня.

Подальший розвиток системи розглядається з позиції розширення функціональних можливостей та підвищення зрілості архітектури, а саме:

- Мультимодальність — обробку аудіо та емоційних сигналів клієнта;
- Інтеграцію з CRM і HR-платформами для персоналізованого аналізу;
- Автоматичне донавчання (fine-tuning) на внутрішніх даних;
- Механізми калібрування й IRR-перевірки для контролю стабільності;
- Розширену безпеку — анонімізацію даних, журналювання дій, шифрування.

Розширення цих напрямів створює передумови для переходу від прототипу до інтегрованої системи управління якістю комунікацій у корпоративному середовищі, у якій система не лише оцінює якість, а й керує процесом розвитку персоналу, стаючи частиною корпоративної екосистеми управління клієнтським досвідом (Customer Experience Management).

Запропонована архітектура поєднує модульність, масштабованість і пояснюваність. Вона відображає системний підхід до управління якістю комунікацій та підтверджує можливість створення інтелектуальної системи контролю, де LLM і RAG працюють у єдиному контурі оцінювання й навчання.

2.5.Інфраструктурні та методологічні вимоги до реалізації системи

Реалізація автоматизованої системи контролю якості комунікацій (АСКЯК) потребує цілісного підходу до інфраструктури, розробки, інтеграції та забезпечення якості. На відміну від класичних інформаційних систем, робота з мовними моделями, RAG-підходом та корпоративними стандартами потребує гарантій безпеки, відтворюваності результатів і стабільності роботи у змінному середовищі контакт-центру. Основні вимоги охоплюють організацію середовища розробки, архітектуру розгортання, принципи модульної інтеграції, механізми валідації, безпеку даних та вимоги до продуктивності, що реалізуються через узгоджену взаємодію ключових інфраструктурних компонентів системи, схематично представлену на рис. 2.6.

Рисунок 2.6. Взаємодія компонентів середовища інфраструктури.

Організація середовища розробки й експлуатації передбачає наявність контрольованого середовища для виконання LLM- та RAG-компонентів, що включає керування залежностями, точні версії моделей, базу знань і структуру промптів. Для забезпечення відтворюваності системи використовуються ізольовані середовища (virtualenv, Conda, Docker), які дозволяють уникати конфліктів між версіями бібліотек і забезпечують контроль над параметрами

моделі. У процесі розробки важливо підтримувати стабільність бази знань, яка включає корпоративні політики, чек-листи, оцінні матриці, приклади діалогів та шаблони промптів. Це забезпечує узгодженість результатів і можливість подальшого масштабування системи.

Практична реалізація автоматизованої системи контролю якості потребує визначення формату її розгортання, оскільки саме інфраструктурна модель задає можливості обчислення, рівень безпеки, вартість підтримки та швидкість реакції системи. У сучасних дослідженнях з побудови LLM-орієнтованих платформ для оцінки діалогів [2][5] виділяють три базові варіанти розміщення: повністю локальне середовище, інтеграція через зовнішній API або гібридна конфігурація.

Оскільки різні варіанти розгортання системи відрізняються не лише архітектурною логікою, але й рівнем контролю над даними, швидкодією, вимогами до ресурсів і гнучкістю адаптації, доцільно спочатку подати узагальнену матрицю їх ключових характеристик. Така матриця дозволяє побачити сильні й слабкі сторони кожного підходу на високому рівні, перш ніж переходити до більш детального порівняння, узагальнене представлення якого наведено на рис. 2.7.

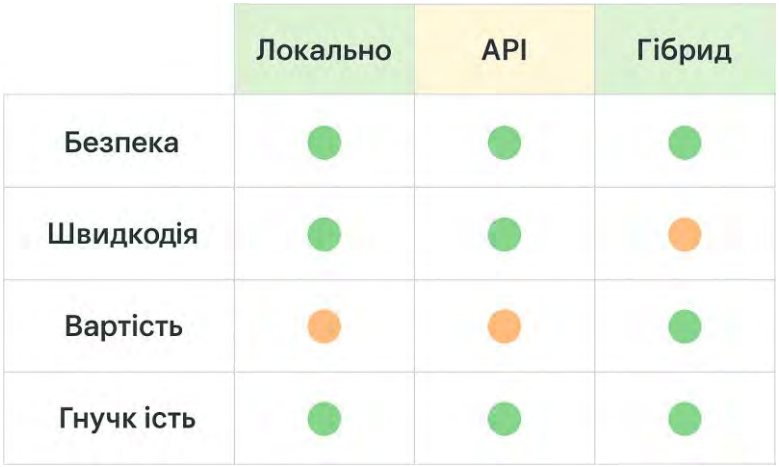

Рисунок 2.7. Матриця ключових характеристик системи.

Після оглядової матриці (див. табл. 2.3) представлено розгорнуту таблицю, яка надає глибшу технічну характеристику трьох варіантів

розгортання та містить детальні пояснення щодо продуктивності, вартості, безпеки й складності підтримки.

Таблиця 2.3. Порівняння варіантів розгортання системи.

Параметр	Локальне розгортання (Self-hosted)	API-орієнтований режим	Гібридна модель
Контроль над даними	Максимальний. Усі дані зберігаються локально. Найкраще для конфіденційних контакт-центрів.	Низький — дані передаються зовнішньому провайдеру.	Середній. Чутливі дані — локально, специфічні запити — API.
Вимоги до апаратури	Високі (CPU/GPU, обсяг RAM). Потрібні потужні сервери.	Мінімальні. Достатньо стабільного інтернет-з'єднання.	Помірні. Основна обробка локально, складні задачі — API.
Швидкість роботи (latency)	Низька затримка (найвища швидкодія), особливо при квантованих моделях	Залежить від мережі. Затримка вища через мережеві виклики.	Балансована затримка: критичні операції локально, інші — через API.
Вартість утримання	Висока (обслуговування серверів, оновлення LLM).	Оплата за API-запити. Змінна вартість, без інфраструктурних витрат.	Середня. Частина задач дешевша локально, частина — API.
Гнучкість модифікацій	Максимальна. Можна донавчати, оптимізувати та кастомізувати моделі.	Обмежена провайдером. Можна змінювати тільки промпти та логіку використання.	Гнучка: локальні модулі можна налаштовувати, API — доповнює можливості.
Безпека	Найвища. Легше забезпечити відповідність GDPR/ISO 27001 (європейський регламент захисту персональних даних, важливий у роботі контакт-центрів).	Найнижча, залежить від стороннього провайдера та політики передачі даних.	Середня. Чутливі блоки — локальні, інше — через API.
Складність розгортання	Висока (DevOps, оптимізація моделей, моніторинг).	Низька. Швидке підключення через REST API.	Середня. Потрібен баланс між локальними і зовнішніми модулями.

Продовження таблиці 2.3. Порівняння варіантів розгортання системи.

Параметр	Локальне розгортання (Self-hosted)	API-орієнтований режим	Гібридна модель
Надійність / стабільність	Висока, якщо інфраструктура оптимізована.	Залежить від стабільності хмарного провайдера.	Комбінована, але більш стійка до збоїв у цілому.
Оптимальний сценарій використання	Великі контакт-центри, підвищена конфіденційність, високі навантаження.	Стартапи, невеликі команди, швидкий запуск MVP.	Середні компанії, гнучкі системи, потреба в балансі швидкості та вартості.

Методологічна частина реалізації системи базується на принципах валідації, калібрування й регулярного тестування. До ключових процедур належать: міжоцінювальна узгодженість (IRR), яка передбачає порівняння автоматичних оцінок із ручними оцінками аудиторів; A/B-тестування промптів; regression testing для RAG-пайплайна; моніторинг prompt drift і data drift, що дозволяє виявити зміну поведінки моделі при оновленні бази знань або зміні контекстів взаємодії. Дослідження [1][3] показують, що наявність чіткої структури плану оцінювання — ієрархії критеріїв і підкритеріїв — є критичною умовою відтворюваності оцінювальних LLM-систем , а практичну реалізацію цього контуру контролю відображає взаємодія основних модулів системи (рис. 2.8).

Рисунок 2.8. Взаємодія модулів системи.

Безпека даних є базовою передумовою для впровадження автоматизованої системи, оскільки контакт-центри працюють з персональними даними клієнтів. Інфраструктура має забезпечувати шифрування даних під час зберігання й передавання в рамках встановлених стандартів (TLS 1.3, AES-256), анонімізацію ідентифікаторів клієнтів, рольову систему керування доступом користувачів до ресурсів системи (RBAC), журналювання дій та політики зберігання даних у відповідності до регламентів [8]. Для гібридних сценаріїв рекомендується передавати до зовнішніх API лише знеособлені частини діалогів, що мінімізує ризики витоку даних.

Продуктивність і масштабованість системи визначаються балансом між точністю LLM-аналітики, часом відповіді та вартістю обчислень. Дослідження [5] показують, що використання квантованих моделей дозволяє знизити чутлівість до 50 % без критичної втрати якості, що є важливим для MVP у компаніях із середніми обчислювальними ресурсами. Для оптимізації часу відповіді застосовуються batch-processing, кешування контекстів, балансування навантаження між CPU та GPU, а також асинхронна обробка через черги повідомлень. Для контролю продуктивності використовуються Service-Level Indicators (SLI) і Service-Level Objectives (SLO , що являють собою відповідно вимірювані показники рівня сервісу та цільові значення їх досягнення, зокрема середній час відповіді, коефіцієнт помилок, відсоток успішних запитів та кореляцію автоматичних оцінок із людськими оцінками (показник точності класифікації понад 0,8 для ключових критеріїв контролю якості). Масштабованість забезпечується контейнеризацією (Docker) та оркестрацією (Kubernetes, Docker Swarm), що дозволяє адаптувати потужності під поточне навантаження.

Ці інфраструктурні та методологічні вимоги формують основу надійності й відповідальності системи, забезпечують стабільність роботи, точність і безпеку даних, а також створюють умови для еволюції MVP-версії системи до повноцінної корпоративної платформи управління якістю комунікацій.

2.6. Теоретичні основи оцінювання ефективності системи

Оцінювання ефективності автоматизованої системи контролю якості комунікацій (АСКЯК) є критичною складовою її функціональної надійності, оскільки воно визначає, наскільки результати роботи моделі узгоджуються з експертними оцінками, корпоративними стандартами та очікуваним поведінковим патерном агента. У сучасних дослідженнях [1][2][5] підкреслюється, що LLM-орієнтовані системи демонструють високу якість аналітики лише тоді, коли їх валідація здійснюється через комплексний набір метрик, що охоплюють точність, стабільність, пояснюваність та поведінкові характеристики взаємодії.

Теоретичну основу оцінювання системи формують класичні метрики якості — Accuracy, Precision, Recall, F1-score = — які застосовуються для вимірювання того, наскільки модель правильно класифікує порушення або виконання критеріїв.

- Accuracy — частка коректно класифікованих діалогів;
- Precision — точність передбачення позитивного класу (правильних позитивів серед усіх передбачених);
- Recall — здатність виявляти всі позитивні випадки;
- F1-score — гармонійне середнє Precision і Recall, що використовується як узагальнений показник ефективності.

У задачах оцінювання діалогів ці метрики використовуються для перевірки коректності роботи специфічних підкритеріїв, наприклад: “наявність привітання”, “з’ясування потреби”, “деескалація”, “конфліктність”, “ввічливість” тощо, що підтверджено у дослідженнях контакт-центрових систем автоматичного моніторингу [2]. Коли система має відтворювати рішення експерта, важливо аналізувати не лише класифікаційні метрики, а й ступінь узгодженості, тобто наскільки оцінки моделі корелюють з оцінками аудиторів, а також врахувати специфіку діалогових сценаріїв та їхню багатокроковість [40]. Для цього застосовуються коефіцієнти кореляції

(Spearman, Pearson) [1], які дозволяють встановити рівень близькості між людськими та автоматичними оцінками.

Для систематизації підходів до оцінювання ефективності автоматизованої системи доцільно згрупувати метрики за характером їх застосування у таблиці 2.4.

Таблиця 2.4. Групи метрик оцінювання ефективності системи.

Група метрик	Характеристика	Приклади показників
Класичні	Відображають точність виявлення порушень	Точність, повнота, збалансованість
Кореляційні	Порівнюють результати системи з експертною оцінкою	Узгодженість з експертом
Якісні та поведінкові	Характеризують практичну придатність системи	Пояснюваність, стабільність, відтворюваність

Окремої уваги потребує міжоцінювальна узгодженість (Inter-Rater Reliability, IRR), яка традиційно використовується в QA-командах для перевірки однорідності оцінок аудиторів. За аналогією, IRR може застосовуватися як метрика порівняння LLM-системи з одним або кількома людськими оцінювачами. Якщо кореляція між автоматичною та експертною оцінкою перевищує 0.75, що у літературі розглядається як прийнятний рівень валідності для базових сценаріїв — система вважається валідною для базових сценаріїв (Jia et al., 2024). Саме IRR і кореляційні метрики утворюють фундаментальний рівень оцінювання якості MVP-версії системи.

Пояснюваність (explainability) є ще одним центральним аспектом ефективності LLM-орієнтованої системи. На відміну від класичних моделей, сучасні мовні моделі здатні аргументувати свої рішення у формі природномовного reasoning, виділяючи фрагменти діалогу, які стали основою для конкретної оцінки. У дослідженнях [6] показано, що моделі, навчені з RLHF, що являє собою метод навчання моделі з підкріпленням на основі людських оцінок, демонструють значно кращу здатність до прозорого аргументування. Пояснюваність прямо корелює з довірою аудиторів до системи. Саме тому в АСКЯК пояснюваність стає не додатковою опцією, а функціональною вимогою: вона дозволяє відслідковувати причинно-

наслідкові зв’язки, забезпечує прозорість оцінювання та підтримує процес коучингу, реалізуючись у вигляді поетапного формування пояснення — від виділення релевантних фрагментів діалогу до генерації обґрунтованих рекомендацій (рис. 2.9).

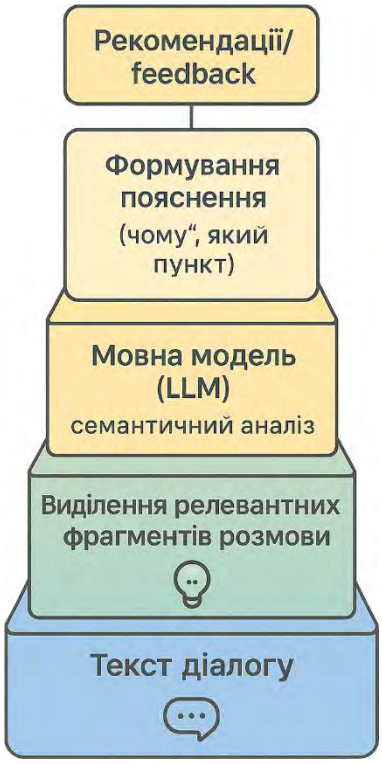

Рисунок 2.9. Формування пояснення як важлива частина рекомендацій.

За даними [3], додавання етапу експертного уточнення після попереднього оцінювання LLM підвищує Macro F1 системи до 0.81 — на ≈ 19 % краще, ніж у повністю автономному режимі. Такий підхід забезпечує не лише підвищення точності, а й покращення навчальної здатності системи: експертні коментарі використовуються як зворотний зв’язок для подальшого fine-tuning або prompt optimization.

Роль людини у циклі оцінювання включає:

- валідацію складних кейсів (коли дані неоднозначні);
- калібрування моделей (порівняння результатів системи з оцінками аудиторів);
- етичний контроль (перевірка відсутності упереджень або конфліктів у висновках).

Як зазначається [3], саме інтеграція human-in-the-loop забезпечує відповідальність і прозорість системи, що є необхідною умовою її етичної легітимності. Таким чином, людський компонент не суперечить автоматизації, а виконує регулюючу та навчальну функцію, забезпечуючи безперервне вдосконалення моделі.

Для комплексного опису ефективності системи у науковій літературі пропонується інтегральна модель із чотирьох компонентів — так звана 4E-Model, яка охоплює Effectiveness, Efficiency, Explainability та Ethicality. Effectiveness визначає точність і узгодженість оцінок; Efficiency — час відповіді й вимоги до обчислювальних ресурсів; Explainability — якість пояснення та здатність моделі обґрунтовувати свої рішення; Ethicality — дотримання конфіденційності даних, уникнення упередженості та відповідність регуляторним вимогам, а їх взаємозв'язок у межах єдиного підходу до оцінювання ефективності системи відображено в інтегральній моделі з чотирьох компонентів (рис. 2.10).

Рисунок 2.10. Інтегральна модель із чотирьох компонентів.

Представлення системи через ці чотири компоненти дозволяє не лише оцінювати якість її роботи, а й вибудовувати дорожню карту її подальшого розвитку.

На завершальному рівні оцінювання ефективності знаходяться операційні метрики, які визначають придатність системи до реального

використання в контакт-центрі: стабільність роботи, середня затримка відповіді, відсоток невдач обробки, наявність fallback-механізмів, чутливість до зміни контекстів (prompt drift, data drift), стабільність reasoning при змінах формулювання промптів. Згідно з результатами дослідження [5], використання квантованих моделей дозволяє зменшити latency без втрати точності, що є важливою перевагою при розгортанні MVP-версії системи на локальних серверах.

Таким чином, оцінювання ефективності АСКЯК є багаторівневим процесом, що поєднує класичні статистичні метрики, кореляційні показники узгодженості, аналіз поведінкових характеристик, перевірку пояснюваності та операційну надійність системи. Така багатовимірна модель дозволяє гарантувати, що система не лише правильно ідентифікує порушення стандартів, а й робить це стабільно, прозоро й у спосіб, який може бути відтворений і перевірений людиною. Для MVP-версії застосовуються базові метрики точності та пояснюваності, тоді як повна система повинна відповідати розширеним вимогам інтегральної моделі 4E, що відкриває можливість довгострокового масштабування й використання в інфраструктурі контакт-центру.

Висновки до розділу 2

У розділі 2 сформовано теоретичну основу для побудови автоматизованої системи контролю якості комунікацій у контакт-центрі. Показано, що поєднання можливостей великих мовних моделей із механізмами доступу до корпоративної бази знань забезпечує семантично обґрунтований аналіз діалогів та підтримує адаптивність системи в умовах зміни нормативів і бізнес-процесів.

Концептуальна модель системи визначає взаємодію модулів збору, обробки, інтелектуального аналізу та формування пояснень, а архітектура MVP реалізує замкнений цикл «вхід — аналіз — оцінка — пояснення». Це

дозволяє перевірити працездатність ключових механізмів та закласти основу для масштабування.

Інфраструктурні вимоги охоплюють вибір режиму розгортання, забезпечення безпеки даних, контроль версій промптів і моделей, а також методологію валідації. Теоретичні засади оцінювання містять поєднання кількісних, кореляційних і якісних показників, доповнених принципами пояснюваності та участі людини в циклі.

Отримані результати формують підґрунтя для подальшого удосконалення системи, що розглядається у розділі 3, де буде запропоновано розширену модель підкритерійного аналізу та оцінено її ефективність на практичних даних.

3. УДОСКОНАЛЕННЯ СИСТЕМИ КОНТРОЛЮ ЯКОСТІ КОМУНІКАЦІЙ НА ОСНОВІ БАГАТОРІВНЕВОЇ МОДЕЛІ ОЦІНЮВАННЯ

3.1.Передумови та обґрунтування необхідності удосконалення

Сучасні контакт-центри функціонують у висококонкурентному середовищі, де якість комунікації стала ключовим чинником утримання клієнтів, формування їхнього досвіду та оптимізації витрат бізнесу. У практиці галузі відбувається активний перехід до автоматизованих систем контролю якості, значна частина яких включає модулі на основі великих мовних моделей (LLM). Згідно з оглядами провідних постачальників аналітики для контакт-центрів, таких як Zendesk та Freshdesk, зростає запит на масштабовану, об’єктивну та відтворювану оцінку розмов, яка дозволяє охоплювати весь масив взаємодій, а не лише вибіркові аудити. Проте навіть за наявності ШІ-складових більшість сучасних рішень залишаються обмеженими у глибині аналізу й часто відтворюють лише спрощену версію процесів ручного аудиту.

Одним з головних викликів є характер інтеграції LLM у бізнес-процеси. Комерційні платформи на кшталт MaestroQA, Evaluagent або Scorebuddy розширили традиційний механізм оцінки — скоркард — додаванням LLM-модуля для автоматичного скорингу. Проте у більшості випадків мовна модель використовується як “надбудова”, що працює за єдиною інструкцією (single prompt) та здійснює узагальнений висновок про відповідність діалогу конкретному критерію скоркарду. Підхід такого типу дозволяє швидко масштабувати оцінювання, однак не забезпечує повноцінного аналізу причин відхилень, не дає можливості виявляти мікропомилки агента та не відтворює логіку живого аудитора. У статті [1] зазначено, що існуючі LLM-функції здебільшого орієнтовані на отримання узагальненого результату без детальної декомпозиції, що значно знижує цінність таких рішень для тренінгових та коучингових процесів.

Другим аспектом проблематики є відсутність підкритерійного аналізу. У реальній роботі фахівців з контролю якості кожна категорія скоркарду складається з набору підкритеріїв, що відображають типи помилок або стандарти поведінки. Як приклад, розглянемо типовий блок Вирішення запиту клієнта (Solving the customer's request), який включає низку конкретних відхилень: нечітка відповідь, неповні рекомендації, ігнорування ключового запиту, непотрібні перенаправлення, неправильна або неповна ескалація, різні типи дезінформації тощо. Наявні LLM-системи уникають проходження цих підкритеріїв по одному, що підтверджується практичним аналізом рішень Zendesk QA та Scorebuddy. Внаслідок цього вихідна оцінка стає бінарною (Валідно/Невалідно) без декомпозиції причин, що саме було порушено. Наукові дані [1] демонструють, що моделі без структурованого формату reasoning мають значно нижчу кореляцію з людськими оцінювачами, оскільки вони генерують відповідь на основі узагальненого контексту, а не шляхом послідовного аналізу конкретних аспектів діалогу.

Третя проблема — відсутність пояснень та доказів (цитат розмов). Людський аудитор обов'язково посилається на репліки агента й клієнта, пояснює причину порушення та формує рекомендацію. Типові LLM-інструменти на ринку цього не роблять. У дослідженні [3] показано, що моделі без попереднього формування структурованого плану оцінювання ("evaluation plan") схильні до нестабільної поведінки та можуть змінювати свої висновки за незначної зміни формулювань інструкції або контексту. Це явище відоме як prompt drift. Наявність пояснень (explainability) — не просто корисна функція, а критичний елемент достовірності, особливо у регульованих доменах, де оцінка має бути аргументованою та відтворюваною, що узгоджується з висновками про ризики недостатнього нагляду в практичних діалогових ШІ-системах [41].

Четвертим недоліком є відсутність персоналізації оцінювання залежно від типу клієнта та складності ситуації. У бізнес-практиці виділяються категорії VIP-клієнтів (клієнти підвищеної цінності, аналіз яких потребує

персоналізації), для яких вимоги до стандартів обслуговування значно вищі. Більшість існуючих платформ не дозволяють змінювати вагу підкритеріїв або порогові значення залежно від важливості звернення, що призводить до вирівнювання оцінок і втрати управлінського сенсу метрики. Аналітичні огляди [9] підкреслюють, що диференціація клієнтських сценаріїв є ключовим чинником підвищення якості сервісу, проте сучасні AI-QA рішення здебільшого позбавлені таких можливостей.

П’ятим структурним обмеженням є брак навчальної придатності. Компанії очікують, що автоматизоване оцінювання не просто поставить оцінку, а й підкаже, які саме компетенції потребують розвитку. У дослідженні [10] зазначено, що саме елемент персоналізованого коучингу є найбільш цінним для агентів і супервайзерів, оскільки дозволяє швидко усувати слабкі місця та підвищувати стандарти сервісу. Проте на практиці LLM-інтеграції у QCS обмежуються висновками високого рівня (“недостатня чіткість”, “неповне вирішення”). Вони не формують операційного фідбеку на рівні: “Порушення: неправильна ескалація. Репліка: .. Причина: .. Як виправити: ..”

Зазначені обмеження свідчать про наявність розриву між потенціалом сучасних мовних моделей та практиками їх використання у системах автоматизованого контролю якості. За відсутності структурованого підходу до оцінювання LLM не відтворюють глибину, послідовність і аргументованість, характерні для роботи живого аудитора. Це зумовлює доцільність розроблення багаторівневої моделі оцінювання, яка передбачає формалізований прохід за підкритеріями, автоматичне формування доказових фрагментів та генерацію коучингового зворотного зв’язку. Концептуальні засади такої моделі розглянуто у підрозділі 3.2.

3.2. Концепція багаторівневої моделі оцінювання як основи удосконаленої системи контролю якості

Автоматизація контролю якості комунікацій у контакт-центрах розвивається у напрямі інтеграції великих мовних моделей, проте більшість

існуючих рішень реалізують аналіз на рівні загальних блоків скоркарду. У таких системах категорія типу «Вирішення запиту клієнта» оцінюється як єдина сутність, що призводить до поверхневого висновку й неможливості визначити конкретні причини порушення. Цей недолік фіксується як у практичних оглядах ринку (Zendesk, Qualtrics, Scorebuddy), так і в наукових роботах. Зокрема, у статті [2] вказується, що LLM-модулі сучасних QA-платформ переважно застосовують глобальні промпти й не забезпечують структурованого аналізу. Наукові дослідження [1][3] показують, що валідність оцінювання суттєво зростає, коли модель працює за структурованим планом і проходить серію підкритеріїв, а не генерує узагальнений висновок.

Удосконалення системи контролю якості, запропоноване у цій роботі, ґрунтується на переході від блокового аналізу до багаторівневої моделі оцінювання. Ця модель відтворює логіку крок-за-кроком, за якою працюють фахівці з якості: спочатку визначається контекст розмови, потім послідовно перевіряються конкретні поведінкові ознаки агента, далі формується доказові фрагменти діалогу, і лише після цього агрегується загальний висновок і рекомендації. Такий підхід розширює можливості LLM і дозволяє отримати обґрунтовані, відтворювані та інтерпретовані результати.

Багаторівнева модель оцінювання — це формалізована схема аналізу діалогу, яка декомпонує загальний критерій скоркарду на низку конкретних підкритеріїв. Кожен підкритерій є мінімальною поведінковою одиницею, яку можна ідентифікувати у діалозі й перевірити на відповідність корпоративним стандартам. Така структура замінює традиційний одноетапний block-level аналіз на покрокову когнітивну процедуру, що нагадує роботу живого аудитора.

Формалізований опис багаторівневої моделі оцінювання

Для забезпечення відтворюваності, автоматизованості та можливості алгоритмічної реалізації багаторівнева модель оцінювання може бути подана у формалізованому вигляді.

Нехай діалог між клієнтом і агентом подано як упорядковану послідовність реплік:

$$D = \{r_1, r_2, \dots, r_n\},$$

де кожна репліка $r_i = (a_i, t_i, x_i)$ характеризується автором $a_i \in \{client, agent\}$, часовою міткою t_i та текстовим вмістом x_i.

Для кожної категорії контролю якості C_k (наприклад, «Вирішення запиту клієнта») визначається множина підкритеріїв:

$$S_{kj} = \{S_{k1}, S_{k2}, \dots, S_{km}\},$$

де кожен підкритерій S_{kj} є формалізованою поведінковою ознакою, що описується кортежем:

$$S_{kj} = \langle \mathrm{def}_{kj}, T_{kj}, E_{kj} \rangle,$$

Де def_{kj} — формальне визначення підкритерію,

T_{kj} — множина типових тригерів або індикаторів,

E_{kj} — очікувана поведінка агента відповідно до корпоративного стандарту.

Оцінювання підкритерію здійснюється автоматизованою функцією аналізу:

$$f_{kj}(D) \to \langle v_{kj}, e_{kj} \rangle,$$

де $v_{kj} \in \{0,1,2\}$ — рівень порушення (0 — відсутнє, 1 — незначне, 2 — критичне),

$e_{kj} \subseteq D$ — множина доказових фрагментів (evidence), тобто реплік діалогу, що підтверджують прийняте рішення.

Функція f_{kj} реалізується мовною моделлю (LLM), яка виконує покроковий reasoning: ідентифікацію релевантних реплік → співставлення зі стандартом → формування локального пояснення → класифікацію рівня порушення. Такий підхід унеможливлює генерацію узагальненого рішення без аналізу конкретних поведінкових ознак.

Для формування інтегральної оцінки категорії використовується агрегувальна функція:

$$Score(C_k) = g\left(\sum_{j=1}^{m} w_{kj} \cdot v_{kj}\right),$$

де w_{kj} — ваговий коефіцієнт підкритерію, що може змінюватися залежно від типу клієнта або сценарію (наприклад, для VIP-звернень),
$g(\cdot)$ — функція нормалізації, яка відображає сумарний бал у фінальний клас оцінювання (OK / Minor / Major).

Формування коучингового зворотного зв'язку описується відображенням:

$$Feedback = h\left(\{v_{kj}, e_{kj}\}_{j=1}^{m}\right),$$

де $h(\cdot)$ — функція генерації рекомендацій, яка на основі матриці підкритеріїв автоматично формує пояснення причин відхилень і практичні поради щодо їх усунення.

Таким чином, багаторівнева модель оцінювання подається як формалізований алгоритмічний процес, у якому всі етапи — від аналізу діалогу до формування фінального QA-висновку — виконуються автоматизованим LLM-модулем без участі людини у прийнятті рішення. Роль експерта обмежується налаштуванням підкритеріїв, вагових коефіцієнтів і стандартів, що підтверджує автоматизований характер запропонованої системи контролю якості.

Дослідження [1] показує, що застосування структурованого reasoning підвищує кореляцію результатів моделі з людськими оцінками. У роботі [3] доведено, що LLM моделі демонструють значно вищу стабільність та меншу чутливість до перестановок інструкцій (промптів), коли працюють не за глобальною інструкцією, а відповідно до попередньо сформованого плану аналізу.

У межах цього дослідження багаторівнева модель визначається як семирівнева структура, що поєднує: категорії, підкритерії, стандарти, тригери, докази, логіку рішень, коучингові рекомендації.

Архітектурна логіка багаторівневої моделі

Основна ідея полягає в тому, що загальний критерій скоркарду (наприклад, «Вирішення запиту клієнта») розбивається на набір чітких поведінкових підкритеріїв. Замість класичного глобального промпта (“Чи відповідає розмова стандартам?”), модель проходить послідовний багатоступінчастий аналіз.

Такий підхід вимагає не лише зміни формулювання запитів до моделі, а й перебудови самої логіки оцінювання — від монолітного рішення до поетапного аналізу з фіксацією проміжних результатів. Саме тому в межах удосконаленої системи контролю якості пропонується багаторівнева модель, у якій кожен підкритерій має власну інструкцію, ознаки та очікувані докази. Для відображення цього процесу наведемо схему, яка демонструє:

- як промпт-інженер системи налаштовує підкритерії;
- як вони перетворюються на окремі інструкції;
- як LLM обробляє кожну інструкцію;
- як формується підсумкова оцінка.

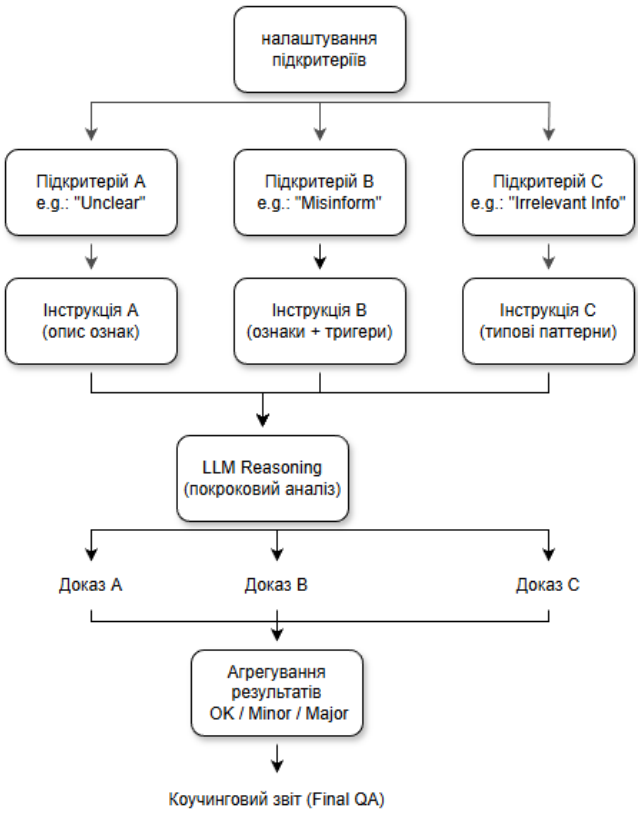

Рисунок 3.1. Процес формування та застосування підкритерійних інструкцій.

Схема відображає ключовий момент: замість одного глобального промпта система генерує й застосовує послідовний ланцюг структурованих інструкцій для кожного підкритерію — від їх налаштування до агрегації результатів у фінальну QA-оцінку. Саме цей механізм підвищує пояснюваність та точність.

Переваги концепції випливають із самої глибини представлення діалогу:

- Глибина аналізу: підкритерії дозволяють виявляти конкретні поведінкові порушення, а не агрегований бал.
- Відтворюваність рішень: чітка структура зменшує нестабільність відповідей моделі при однаковому завданні.
- Пояснення: докази є обов'язковою частиною покрокового аналізу.
- Підтримка VIP-сегментів: різні вагові коефіцієнти.
- Сумісність з CoT-підходами, що являють собою генерацію відповідей, що включає покрокове логічне міркування моделі. [11]
- Підтримка reasoning-планів [12]
- Відповідність індустріальним рекомендаціям [13]

Щоб продемонструвати практичний механізм роботи багаторівневої моделі оцінювання, розглянемо одну з найважливіших категорій у системах контролю якості — «Вирішення запиту клієнта», яка безпосередньо відображає здатність агента не лише надати відповідь, а й довести діалог до результативного завершення. У традиційних QA-рішеннях (зокрема MaestroQA-подібних) ця категорія зазвичай фіксується одним агрегованим балом, тоді як у запропонованому підході вона декомпонується на формалізовані підкритерії, що відображають окремі атомарні поведінкові відхилення агента та створюють основу для доказового аналізу.

Для демонстрації того, як багаторівнева модель реалізується на рівні формалізованих інструкцій і тригерів, у таблиці 3.1 наведено приклад декомпозиції однієї з ключових категорій скоркарду.

Для кожного підкритерію подані:

1. формальне визначення;

2. типовий тригер або лінгвістичний патерн;

3. очікувана поведінка агента згідно корпоративного стандарту.

Таблиця 3.1 — Приклад багаторівневої моделі оцінювання категорії Вирішення запиту клієнта (Solving the customer's request).

Підкритерій	Формальне визначення	Типові тригери / індикатори	Очікувана поведінка агента
Unclear or complicated response	Відповідь є нечіткою, надто загальною або заплутаною, що не допомагає клієнту зрозуміти наступні кроки.	«Я не знаю», «Можливо», «Спробуйте самостійно…», нечіткі або розмиті формулювання.	Надавати структуровану, однозначну та зрозумілу відповідь із чіткими кроками.
Incomplete resolution	Агент дає часткове або неповне рішення, пропускаючи один чи кілька важливих етапів.	Відсутність уточнень, відсутність перевірки стану послуги, неповне пояснення процедури.	Повністю завершити дію: перевірити, уточнити, надати повний алгоритм вирішення.
Irrelevant or redundant information	Надання інформації, не пов'язаної із запитом, або повторення несуттєвих деталей.	Розлогі пояснення, не пов'язані з проблемою; «перевантажені» відповіді.	Фокусуватися на релевантній інформації, уникати надлишкових пояснень.
Lack of probing / уточнювальних запитань	Агент не зібрав ключову інформацію, необхідну для вирішення запиту.	Відсутність запитань типу «Уточніть…», «Підкажіть, будь ласка…».	Поставити мінімальний набір уточнень для точного визначення проблеми.
Untimely response	Агент надто довго затримує відповідь або не реагує оперативно.	Пауза без пояснення, затримка понад регламент.	Тримати клієнта в курсі, пояснювати паузи, відповідати в межах стандартів SLA.
Canned response abuse	Використання шаблонних або повторюваних відповідей без адаптації до контексту.	Повтор однієї й тієї ж фрази, відсутність персоналізації.	Пояснення має бути адаптованим під конкретний випадок, а не шаблонним.
Minor misinformation	Незначна неточність, яка не змінює суті, але може вводити клієнта в оману.	Некоректні формулювання, неправильні деталі другорядного значення.	Надавати повністю точну інформацію, перевіряти дані перед відповіддю.
Major misinformation	Надання неправильної або суперечливої офіційним даним інформації, що впливає на результат.	Неправильні дані про тариф, політику, умови, процедури.	Надавати лише підтверджену інформацію, уникати припущень.

Продовження таблиці 3.1 — Приклад багаторівневої моделі оцінювання категорії Вирішення запиту клієнта (Solving the customer’s request).

Підкритерій	Формальне визначення	Типові тригери / індикатори	Очікувана поведінка агента
No escalation when required	Агент не ескалував випадок, хоча цього вимагає політика чи логіка ситуації.	«Я не можу допомогти, перевірте на сайті», відмова передати звернення.	Ескалувати згідно з регламентом у ситуаціях, що перевищують рівень компетенцій агента.
No alternatives provided	Агент не запропонував клієнту варіантів дії у випадку, коли основне рішення недоступне.	«Це все», «Я нічого не можу запропонувати».	Надати мінімум два варіанти дій (альтернативи).

Подана таблиця ілюструє, як загальна категорія перетворюється на набір формалізованих поведінкових елементів, що безпосередньо інтегруються в підкритерійні інструкції та reasoning-ланцюги LLM. Такий підхід забезпечує:

- прозорість логіки оцінювання,
- чітку структуру reasoning для LLM,
- можливість формування доказів (evidence),
- міцний фундамент для автоматизованих коучингових рекомендацій,
- точну діагностику якості роботи агента, що недоступно у традиційних block-level системах.

Багаторівнева модель відіграє роль когнітивного ядра, яке задає когнітивне ядро всієї системи оцінювання. Вона:

- визначає, як саме LLM має мислити;
- керує послідовністю кроків reasoning;
- визначає точну форму evidence;
- формує основу для автоматизованого коучингу.

У роботі [5] підкреслюється, що системи з багатоступеневою логікою прийняття рішень забезпечують кращу стабільність і гнучкіше масштабування.

Узагальнюючи, можемо зазначити, що така модель оцінювання формує підґрунтя для переходу від традиційного блокового оцінювання до глибокої, доказової та відтворюваної аналітики діалогів. Вона дозволяє LLM працювати

не як генератор загальних висновків, а як когнітивний аудитор, що послідовно перевіряє підкритерії, формує доказові фрагменти діалогу / цитати та надає персональні коучингові рекомендації. Це створює фундамент для архітектури удосконаленого LLM-модуля.

3.3. Архітектура удосконаленого LLM-модуля підкритерійного аналізу

Удосконалена система контролю якості комунікацій передбачає введення спеціального LLM-модуля, що працює за принципами структурованого когнітивного аналізу та забезпечує глибоку оцінку діалогів відповідно до багаторівневої моделі. На відміну від традиційних рішень, де мовна модель виконує одноразовий блоковий скоринг, запропонований модуль здійснює поетапне осмислення діалогу, що значно підвищує точність, стабільність та відтворюваність результатів. В основі архітектури лежить ідея, що LLM розглядається як когнітивний аудитор, що послідовно аналізує поведінкові сигнали агента, співставляє їх зі стандартами компанії, формує обґрунтування своїх рішень та надає рекомендації.

Дослідження [3][1] показують, що ступінь валідності результатів LLM значно зростає, коли модель працює відповідно до заздалегідь визначеної когнітивної схеми. У такому випадку вона не “вгадує” відповідь на основі загального контексту, а проходить структурований план аналізу, що робить результати ближчими до рішень експертів. На цій концепції побудовано запропонований удосконалений LLM-модуль, архітектура та функціональні компоненти якого розглядаються у підрозділах 3.3.1–3.3.6.

Принциповою особливістю запропонованого підходу є те, що удосконалений LLM-модуль функціонує як повністю автоматизована підсистема контролю якості. Усі етапи — від аналізу транскрипту діалогу до формування підсумкової оцінки та коучингових рекомендацій — виконуються без участі людини в контурі прийняття рішень. Роль експерта обмежується попереднім налаштуванням підкритеріїв, стандартів і вагових коефіцієнтів, що

дозволяє розглядати систему саме як автоматизовану систему контролю якості комунікацій, а не лише аналітичну або рекомендаційну платформу.

3.3.1. Загальна концепція удосконаленого LLM-модуля

Удосконалений модуль поєднує багаторівневу модель оцінювання зі структурованим reasoning мовної моделі. Його основна ідея полягає в тому, щоб замінити одноразову генерацію результату на багатокроковий процес: інтерпретація діалогу → застосування підкритеріїв → пошук доказів → класифікація → узагальнення → коучинг. Така послідовність дає змогу досягти ефекту імітації експертного аудиту, тобто максимально наблизити роботу LLM до логіки живого спеціаліста з якості.

Принципово важливо, що модуль працює у режимі розподіленої відповідальності. Збір знань здійснюється через RAG-компонент, оцінювання — через когнітивне ядро reasoning, а формування рекомендацій — через окремий коучинговий блок. Поділ функцій відповідає тенденціям [5], де підкреслюється необхідність модульності та розмежування ролей між компонентами системи.

3.3.2. Загальна схема обробки діалогу

На початковому етапі модуль отримує транскрипт діалогу, метадані клієнта та службову інформацію. Далі текст очищується, нормалізується й готується до семантичної інтерпретації. Після цього система звертається до корпоративної бази знань, витягаючи релевантні політики, правила, внутрішні процедури чи довідкові матеріали. Це забезпечує фактичну коректність аналізу — критичний аспект, оскільки моделі без RAG-супроводу мають тенденцію до генерації неточностей.

Важливо зазначити, що описана послідовність обробки реалізується в автоматизованому режимі для кожного діалогу, що надходить у систему. Модуль не потребує ручного запуску або втручання аудитора: транскрипти обробляються пакетно або в потоковому режимі, що забезпечує повне

охоплення діалогів і усуває вибірковість, характерну для ручного контролю якості.

Після отримання релевантних знань у роботу вступає блок завантаження моделі оцінювання. На цьому етапі система підбирає дерево підкритеріїв, яке має бути застосоване до конкретної категорії, а також визначає вагові коефіцієнти для кожного критерію, зокрема для VIP-сценаріїв.

Основна робота виконується в ядрі покрокового аналізу, де модель послідовно застосовує компоненти багаторівневого плану до транскрипту. По завершенні аналізу формується матриця підкритеріїв, а на її основі — інтегральну оцінку категорії та коучингові рекомендації.

3.3.3. Модуль підкритерійного покрокового аналізу

Модуль reasoning — центральний компонент удосконаленої системи. Він формалізує логіку, за якою модель інтерпретує поведінку агента, і забезпечує відтворюваність аналізу. На відміну від традиційної генерації відповіді, де вся оцінка залежить від єдиного промпта, підкритерійний reasoning працює як послідовна процедура, що зменшує варіативність і підвищує стійкість результатів.

Робота модуля складається з кількох етапів. Спершу модель узагальнює контекст діалогу, визначаючи ключові потреби клієнта, реакції агента та структуру звернення. Наступний крок — проходження дерева підкритеріїв. Для кожного підкритерію LLM аналізує відповідні репліки, порівнює їх зі стандартом і формує локальне пояснення. Така локальна інтерпретація дозволяє уникнути ситуацій, коли модель формує висновок на основі поверхневих сигналів чи випадкових елементів діалогу.

Після аналізу модель визначає рівень порушення: відсутнє (OK), незначне (Minor) або критичне (Major). У процесі reasoning формується evidence — конкретні фрагменти діалогу, що підтверджують висновок. Це підвищує прозорість оцінювання та відповідає вимогам explainability,

описаним у роботах [3][2]. Завершальний етап — синтез локальних оцінок у глобальний висновок, який визначає якість виконання категорії загалом.

3.3.4. Інтеграція багаторівневої моделі з RAG-архітектурою

Система використовує RAG як механізм підвищення фактичної достовірності. Це особливо важливо для підкритеріїв, пов'язаних із достовірністю інформації (misinformation), дотриманням політик або процедур. Поєднання retrieval та reasoning забезпечує, що модель порівнює репліки агента не лише із шаблонними очікуваннями, а й з реальними правилами роботи компанії.

Коли LLM аналізує, наприклад, підкритерій ескалації, вона отримує з бази знань регламент, що визначає, у яких ситуаціях ескалація обов'язкова, а в яких — факультативна. Це дозволяє уникнути хибної класифікації та помилкових висновків. Аналогічним чином працює підкритерій misinformation: модель співставляє слова агента з офіційними даними, що дає змогу виявляти як незначні неточності, так і серйозні фактологічні помилки.

3.3.5. Модуль класифікації та агрегування результатів

Після завершення reasoning формується матриця підкритеріїв. Кожен рядок містить статус підкритерію, пояснення та evidence. Ця матриця перетворюється на інтегральну оцінку за категорією. Ключовою особливістю є можливість налаштовувати ваги підкритеріїв. Наприклад, для VIP-обслуговування підкритерії, пов'язані з misinformation або відсутністю ескалації, можуть мати вищий коефіцієнт впливу.

Такий підхід забезпечує не лише гнучкість, а й відповідність бізнес-реальності, де різні сегменти клієнтів потребують різної суворості аудиту. На відміну від традиційних систем, де всі категорії та підкритерії рівнозначні, удосконалений модуль дозволяє адаптувати оцінювання відповідно до бізнес-пріоритетів. Таким чином, агрегування виконує не лише підсумкову, а й керуючу функцію в автоматизованій системі оцінювання, оскільки саме на

цьому етапі без участі людини формується фінальний клас якості, який надалі використовується для аналітики, коучингу та управлінських рішень.

3.3.6. Модуль формування коучингового висновку

Останній компонент архітектури — генератор рекомендацій. Він використовує матрицю підкритеріїв і формує деталізований звіт, який містить конкретні порушення, пояснення та поради. Такий формат відповідає сучасним трендам у контакт-центрах, де контроль якості розглядається як частина навчального циклу. Згідно з дослідженнями [10], персоналізований фідбек з evidence значно підвищує ефективність розвитку агентів і зменшує кількість повторних помилок.

Таким чином, описані у підрозділах 3.3.1–3.3.6 модулі утворюють єдину логічно узгоджену архітектуру удосконаленої системи контролю якості комунікацій. Для узагальнення взаємодії компонентів, потоків даних, механізмів керування критеріями та зворотних зв’язків на рис. 3.2 наведено загальну архітектурну схему автоматизованої системи, що реалізує багаторівневу модель оцінювання на основі мовних моделей.

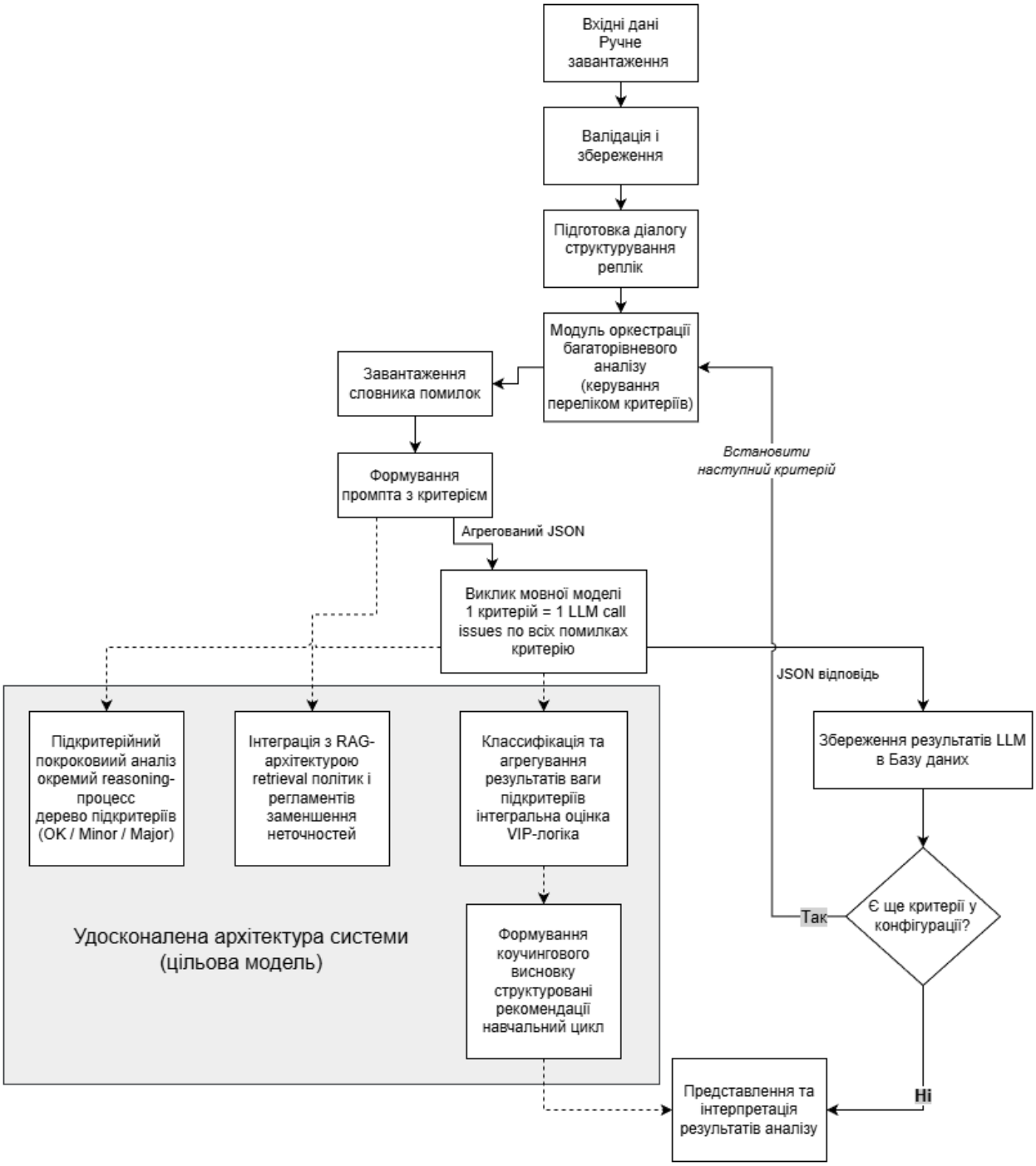

Рисунок 3.2. Узагальнена архітектура удосконаленої системи контролю якості комунікацій.

Наведена схема слугує концептуальним підґрунтям для подальшої експериментальної перевірки core-частини системи, результати якої розглянуто в наступному розділі.

3.4.Генерація автоматизованого коучингового зворотного зв'язку та його роль у системі контролю якості

3.4.1. Значення у процесі забезпечення якості

Контроль якості у контакт-центрах не обмежується фіксацією порушень, оскільки основною метою є підвищення рівня обслуговування та професійного розвитку агентів. Коучинговий зворотний зв'язок є завершальним етапом циклу «оцінювання – аналіз – покращення» і виконує ключову функцію перетворення отриманої інформації на практичні дії. Дослідження і показники [10][14][22] свідчать, що системи, інтегровані з персоналізованими рекомендаціями, підвищують продуктивність агентів та зменшують кількість повторюваних помилок. З огляду на це, автоматизація коучингового етапу є не лише бажаною, а й необхідною для ефективного функціонування сучасних QA-систем.

Традиційні MaestroQA-подібні рішення, попри функції автоматичного скорингу, пропонують здебільшого загальні коментарі без структурованої аргументації та без прив'язки до конкретних реплік. Це знижує якість навчання, оскільки агент не отримує чіткої відповіді на запитання: які конкретні дії агента не відповідали стандартам і якою має бути коректна модель поведінки у аналогічних ситуаціях надалі.

3.4.2. Формування автоматизованих рекомендацій у запропонованій системі

Удосконалена система використовує матрицю підкритеріїв, сформовану reasoning-модулем, як основу для генерації структурованого coaching feedback. Для кожного порушення система визначає коротке формулювання, виявляє відповідний фрагмент діалогу (evidence), пояснює природу відхилення та формує рекомендацію відповідно до корпоративних стандартів. Такий формат дозволяє уникнути загальних формулювань і забезпечує чіткі, практичні поради, що безпосередньо відповідають виявленим помилкам.

Коучинговий звіт має чітку структуру: інтегральна оцінка категорії, перелік підкритеріїв, де зафіксовані відхилення, та короткі рекомендації для їх усунення. Деталізація кожного підкритерію включає цитування проблемної репліки та пояснення, чому вона суперечить стандартам обслуговування. Наприклад, у разі підкритерію Incomplete resolution система зазначає, що агент надав часткову відповідь або не зібрав необхідних уточнень, а у рекомендаціях пропонує алгоритм правильних дій.

Для ілюстрації механізму формування коучингового зворотного зв'язку у межах підкритерію Incomplete resolution наведемо короткий приклад розмови у рис.3.3:

- Клієнт: «Я оплатив тариф, але він не активувався».
- Агент: «Спробуйте ще раз перевірити на сайті».

Рисунок 3.3. Приклад розмови, де відсутні уточнення від агента.

У цьому випадку система фіксує підкритерій Incomplete resolution. Evidence містить репліку агента, де відсутні уточнення або перевірка статусу операції. Рекомендація формулюється так: перед перенаправленням необхідно перевірити внутрішній статус транзакції, уточнити деталі та надати клієнту покрокове вирішення. Цей мінімальний приклад демонструє структуру коучингового блоку та його практичну орієнтацію.

Автоматизовані рекомендації також можуть адаптуватися до рівня агента або сегмента клієнта, зокрема підвищувати суворість вимог для VIP-обслуговування. Така гнучкість дозволяє враховувати специфіку різних типів взаємодій та забезпечує індивідуалізований підхід до навчання.

Автоматизований коучинговий зворотний зв'язок у запропонованій системі реалізується як функціональний етап багаторівневої моделі оцінювання, що безпосередньо спирається на результати підкритерійного аналізу та сформовану доказову базу. Така інтеграція дозволяє забезпечити структурованість і однозначність рекомендацій, а також їх відповідність

виявленим порушенням без використання узагальнених або шаблонних формулювань.

3.4.3. Очікувані результати від впровадження удосконаленої моделі оцінювання (з інтегрованим порівнянням “до/після”)

Запровадження багаторівневої моделі оцінювання та удосконаленого LLM-модуля створює умови для суттєвого підвищення точності, стабільності й відтворюваності процесів контролю якості комунікацій. На відміну від традиційних MaestroQA-подібних систем, що функціонують на основі узагальнених інструкцій і блокового скорингу, запропонована архітектура реалізує структуроване багатокрокове оцінювання, орієнтоване на підкритерії. Це дозволяє досягти вищої кореляції з експертними оцінками та надати більш значущі інструменти для розвитку персоналу. У згаданих роботах [1][3] наголошується, що застосування структурованого plan-based reasoning підвищує точність та надійність оцінювання, а також зменшує варіативність рішень, пов’язану з генеративною природою мовних моделей. Ці факти підтверджують ефективність обраної концепції.

Загальні очікувані результати можна умовно поділити на три групи: покращення на рівні моделі, вплив на якість роботи агентів та бізнесові ефекти. На рівні моделі прогнозується підвищення точності та стабільності завдяки структурованому застосуванню підкритеріїв, evidence-mapping та інтеграції з корпоративною базою знань. Це має зменшити кількість хибнопозитивних і хибнонегативних рішень, оскільки система працює не з загальним контекстом, а з формально визначеними критеріями. Крім того, багаторівнева модель оцінювання дає змогу адаптувати логіку аналізу залежно від сегмента клієнта (наприклад, для VIP-обслуговування), що забезпечує суворіший контроль і мінімізує ризики помилок у high-stakes взаємодіях. Підкритерійний підхід також сприяє підвищенню explainability: кожне рішення супроводжується доказовою базою, що полегшує верифікацію та внутрішній аудит.

Очікувані зміни на рівні агентів пов’язані насамперед із якістю коучингового зворотного зв’язку. Удосконалена система формує персоналізовані рекомендації, ґрунтовані на фактичних фрагментах діалогу, а не загальних формулюваннях. Це забезпечує точкове усунення помилок, швидше набуття необхідних навичок та суттєве зменшення повторюваних відхилень. Дослідження [10] демонструє, що такі підходи можуть підвищити ефективність навчання на 12–18%, що особливо важливо для великих контакт-центрів з високою ротацією персоналу. Крім того, автоматизований коучинг сприяє уніфікації стандартів, оскільки рекомендації не залежать від стилю чи досвіду окремого аудитора.

З погляду бізнесових результатів, запропонована система дає змогу суттєво оптимізувати роботу відділів контролю якості. По-перше, збільшується охоплення аудиту до 100%, що раніше було обмежено лише вибірковою кількістю діалогів через обмежені людські ресурси. По-друге, знижується навантаження на QCS-фахівців, оскільки рутинні операції з перевірки, доказування та формування рекомендацій переходять до автоматизованого модуля. По-третє, підвищується якість взаємодії з VIP-клієнтами завдяки адаптивному контролю суворості та точності аналізу. Нарешті, систематичний коучинг і підвищення відповідності стандартам сприяють покращенню ключових бізнес-показників, таких як CSAT (показник задоволеності), CES (показник зусиль клієнта для отримання сервісу) та retention rate.

Різницю між поточними можливостями MaestroQA-подібних систем і удосконаленою моделлю можна стисло продемонструвати у порівняльній таблиці 3.2.

Таблиця 3.2. Порівняльна інформація до і після удосконалення системи.

	ДО	ПІСЛЯ
Рівень аналізу	Блоковий	Підкритерійний
Explainability	Низька	Висока, evidence-based
Стабільність	Варіативна	Відтворювана

Продовження таблиці 3.2. Порівняльна інформація до і після удосконалення системи.

	ДО	ПІСЛЯ
Reasoning	Однокроковий	Багатокроковий, структурований
Coaching	Загальний	Персоналізований
Підтримка VIP	Мінімальна	Адаптивні ваги критеріїв
Охоплення	Вибіркове	100% діалогів

Таке порівняння підкреслює, що основні якісні зміни полягають у переході від поверхневого блокового аналізу до глибокої когнітивної моделі, яка забезпечує стабільність, прозорість та навчальну цінність. Подібні висновки містяться і в дослідженні [2], де зазначено, що майбутнє QA-систем полягає саме у granular-рівні оцінювання з формалізованими принципами reasoning.

Для узагальнення логіки впливу запропонованої моделі на процес контролю якості подамо її у вигляді послідовної схеми на рис.3.3. Вона демонструє шлях від вихідних даних до управлінських та організаційних результатів удосконаленої системи.

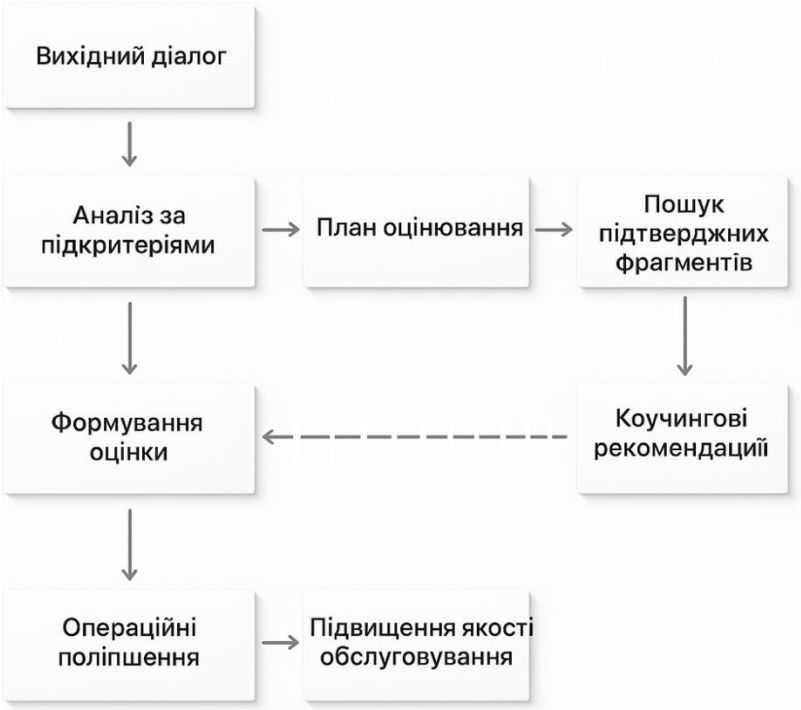

Рисунок 3.4. Послідовна схема від вихідних даних до результату.

Першим етапом є аналіз за підкритеріями, який забезпечує деталізоване опрацювання діалогу на рівні окремих поведінкових ознак. Далі модуль

формує план оцінювання та здійснює пошук підтвердних фрагментів (evidence), що дозволяє обґрунтувати кожне рішення та забезпечити прозорість результатів. Отримані дані передаються у модуль формування оцінки, де відбувається агрегування результатів та визначення рівня відповідності стандартам.

Потік від блоку «Коучингові рекомендації» до блоку «Формування оцінки» позначає зворотний вплив рекомендацій на майбутні взаємодії агентів. На відміну від послідовність основного технологічного процесу, цей потік виділяє поворотний, навчальний зв’язок, який не змінює результат поточного оцінювання, але впливає на якість наступних діалогів. Це означає, що коучингові рекомендації не лише завершують процес аналізу, а й формують підґрунтя для зростання компетентності агентів, што сприяє поступовому зниженню кількості порушень, покращенню взаємодії з клієнтами та підвищенню загальної якості обслуговування.

Описана схема дозволяє формалізувати очікуваний вплив удосконаленої моделі на процес контролю якості та слугує аналітичним підґрунтям для подальшої експериментальної перевірки, результати якої наведено в наступному розділі.

Висновки до розділу 3

У третьому розділі обґрунтовано доцільність переходу від фрагментарних bolt-on AI-рішень до цілісної архітектури автоматизованого контролю якості комунікацій, у якій мовна модель використовується як центральний аналітичний компонент. Показано, що традиційні підходи, зокрема системи типу MaestroQA, забезпечують масштабування аудиту, однак залишаються обмеженими за глибиною аналізу та не реалізують структурованого підкритерійного розбору дій агента.

Запропоноване удосконалення базується на багаторівневій моделі оцінювання, у якій кожна категорія якості деталізується через набір підкритеріїв із формалізованою логікою перевірки. Такий підхід дозволяє

мовній моделі відтворювати reasoning, наближений до логіки експертного аудитора, підвищує відтворюваність результатів та зменшує залежність оцінювання від універсальних промптів. Інтеграція LLM-модуля з RAG-архітектурою забезпечує зіставлення реплік агента з корпоративними політиками та стандартами, що підвищує фактичну коректність аналізу та знижує ризик хибних висновків.

Окремо визначено роль автоматизованого коучингового зворотного зв’язку як складової системи управління якістю, що перетворює результати контролю на практичні рекомендації для розвитку компетентностей агентів. Запропонована архітектура формує методологічну основу для побудови MVP та подальшої експериментальної перевірки ефективності автоматизованого підкритерійного аналізу, що реалізується в наступному розділі роботи.

4. ЕКСПЕРИМЕНТАЛЬНА ЧАСТИНА

Експериментальна частина роботи логічно продовжує теоретичні та методологічні результати, отримані в попередніх розділах, і має на меті перевірити практичну реалізованість автоматизованої системи контролю якості комунікацій у контакт-центрі. У розділах 2 та 3 було сформовано концептуальну основу дослідження: визначено критерії якості розмов, формалізовано перелік типових помилок агентів, побудовано архітектуру MVP-системи та описано алгоритмічний підхід до взаємодії з LLM, за якого виявлення відхилень від стандартів обслуговування здійснюється без участі людини в контурі оцінювання.

Теоретичні висновки свідчать, що застосування LLM у багаторівневій моделі оцінювання може підвищити відтворюваність і точність аудитів за умови, що модель функціонує в межах чітко формалізованих правил, структурованих словників помилок і наперед визначеної логіки reasoning. Водночас ці твердження потребують емпіричного підтвердження, оскільки реальні діалоги характеризуються високою варіативністю, неоднозначними формулюваннями, фрагментарним контекстом та різними сценаріями поведінки агента і клієнта, що створює додаткові виклики саме для автоматизованого аналізу.

Саме тому проведення експерименту стало необхідним етапом роботи:

- щоб перевірити, наскільки MVP-система в автоматичному режимі відтворює логіку експертного аудиту;
- визначити стабільність відповідей моделі при аналізі діалогів різної структури без ручної корекції результатів;
- оцінити практичну цінність структурованих словників помилок як основи автоматизованого контролю;
- встановити реалістичні межі застосування моделі gpt-4.1-mini у задачах повністю автоматизованого контролю якості комунікацій.

Експеримент також дозволяє з’ясувати, які елементи архітектури MVP функціонують ефективно в автоматичному контурі, а які потребують подальшої оптимізації — зокрема, деталізація описів помилок у базі даних, формування промптів, обробка JSON-відповідей (текстовий формат обміну даними між компонентами системи) моделі та механізми інтерпретації evidence-фрагментів без залучення аудитора.

Таким чином, експериментальна частина має не лише підтверджувальний, а й діагностичний характер: вона дає змогу оцінити коректність функціонування автоматизованої системи контролю якості як цілісного програмно-алгоритмічного рішення, визначити її сильні та слабкі сторони, а також сформувати висновки щодо можливостей подальшого масштабування та промислового використання проєкту.

4.1. Визначення та опис умов проведення експерименту

Предметом дослідження в даному експерименті є здатність розробленої MVP-системи автоматизованого контролю якості комунікацій коректно і стабільно виявляти типові помилки у діалогах між клієнтом і агентом контакт-центру відповідно до формалізованих критеріїв оцінювання. Іншими словами, дослідження спрямоване на перевірку того, перевірку того, як автоматизований алгоритм контролю якості, у якому модель gpt-4.1-mini виступає обчислювальним модулем, інтерпретує реальні фрагменти розмов, порівнює їх з еталонними правилами, структурованими у базі даних, та повертає операціоналізовану оцінку у вигляді стандартизованого JSON.

Об’єктом аналізу є п’ять діалогів, представлених у форматі CSV (табличний формат зберігання даних), що відображають різні сценарії взаємодії з клієнтами — від простих інформаційних звернень до ситуацій, у яких можливі помилки в таймінгу, логіці відповідей чи у виявленні потреб користувача. Усі діалоги подано у додатках у повному вигляді, а в цьому розділі вони використані лише як емпіричний матеріал для оцінювання роботи системи.

Експеримент спрямований на перевірку коректності роботи MVP за двома критеріями, які були сформовані та реалізовані на етапі розробки. Повні словники помилок, що визначають зміст кожного критерію, наведено у відповідних таблицях бази даних (див. Додатки А–В). У межах експерименту аналіз здійснювався за такими критеріями:

1. Dialogue Structure & Timings — порушення логіки побудови діалогу, неправильні паузи, невірна черговість або відсутність обов'язкових елементів структури.
2. Identification of the Customer's Needs — помилки у виявленні, уточненні та інтерпретації потреб клієнта, зокрема зайві питання або ігнорування ключових уточнень.

Для кожного діалогу система автоматично запускає повний цикл аналізу за кожним критерієм окремо, без ручного втручання, із формуванням структурованого результату оцінювання.

Умови експерименту визначаються використанням фіксованого середовища, у якому працює MVP-система:
Ubuntu 22.04, веб-сервер Nginx, PHP 8.3, MySQL 8.0.35, панель управління aaPanel та реалізована в проєкті логіка обробки діалогів і виклику LLM. Усі виклики моделі виконувалися через Chat Completions API з використанням gpt-4.1-mini, а параметри запиту були фіксованими для забезпечення повторюваності результатів.

Функціонування довідників помилок — ключова складова умов експерименту. Опис кожної помилки зберігається в базі даних у вигляді стисленого, але точного тексту, який входить до інструкції, що передається моделі. Таким чином, модель не «вгадує» помилки, а працює в рамках чітко заданих правил, які визначають, що саме вважається відхиленням від стандартів. Це дозволяє досягти однакових умов аналізу для всіх діалогів і забезпечує контрольоване середовище для порівняння з результатами ручної експертної оцінки.

Проведення експерименту також передбачало забезпечення сталих параметрів введення: діалоги були попередньо нормалізовані (уніфіковано формат дат, ролей, очищено текст від службових символів), що дозволяє виключити вплив шумів і підвищити узгодженість обробки. Кожен діалог аналізувався окремо за кожним критерієм з формуванням двох незалежних інструкцій і двох відповідей JSON, які зберігалися у таблиці evaluation_runs.

Таким чином, експеримент реалізує замкнений автоматизований контур контролю якості, у якому вхідні дані, правила оцінювання, алгоритм аналізу та формат вихідних результатів є повністю формалізованими й відтворюваними.

Отже, умови експерименту можна охарактеризувати так:

- стандартизовані вхідні дані (до 5 діалогів);
- фіксовані критерії оцінювання й структуровані словники помилок;
- стабільне серверне та програмне середовище;
- єдиний формат виклику LLM та суворий формат JSON-відповіді;
- можливість порівняння результатів із незалежною експертною оцінкою.

Таке визначення предмета та умов дає змогу забезпечити валідність отриманих результатів і переводить експеримент із демонстраційного у справді дослідницький формат, що відповідає вимогам технічної магістерської роботи.

4.2. Вирішення задач підготовки експерименту

Підготовка експерименту вимагала розв’язання низки взаємопов’язаних технічних та методологічних задач, необхідних для забезпечення коректності, повторюваності та валідності отриманих результатів. Основними задачами, що потребували вирішення, були:

1. Уніфікація та підготовка вхідних даних — упорядкування діалогів, очищення тексту, нормалізація форматів та приведення інформації до структури, придатної для автоматичного аналізу.
2. Формування та структуризація словників помилок — підготовка машинно-читаних описів критеріїв і помилок у таблицях бази даних, які будуть використані під час обробки діалогів.

3. Підготовка механізму побудови запитів до моделі — створення інструкцій для формування стабільних запитів, що включають текст діалогу та відповідний набір помилок.
4. Визначення алгоритму експериментального процесу — встановлення чіткої послідовності етапів обробки діалогу: від завантаження CSV до отримання структурованого результату від моделі.
5. Забезпечення відтворюваності експерименту — фіксація параметрів моделі, умов роботи серверного середовища та способу збереження результатів.
6. Організація збереження результатів оцінювання — створення механізму для реєстрації JSON-відповідей у окремій таблиці бази даних з подальшою можливістю аналізу.

Вирішення цих задач створило єдину та стабільну платформу для виконання експериментальної частини роботи. Усі перелічені задачі реалізовані у вигляді автоматизованого підготовчого контуру, що не потребує ручного втручання під час обробки кожного окремого діалогу.

Уніфікація вхідних даних полягала в приведенні діалогів до єдиного формату CSV, який дозволяє автоматично обробляти повідомлення та зберігати їх у таблицях бази даних. Для цього були нормалізовані дати, ролі учасників та текстові фрагменти. Формат CSV, який використовується у системі, наведено в таблиці 4.1.

Таблиця 4.1 — Формат CSV-файлу діалогу.

Поле	Опис
date	Дата і час повідомлення у форматі РРРР-ММ-ДД ГГ:ХХ:СС.мс, що відповідає стандартам журналювання у базі даних.
author	Автор повідомлення. Може набувати значень *agent* або *client*. Системні службові змінні не включаються.
text	Текст повідомлення, який був реально надісланий клієнтом або агентом. Приватні нотатки та внутрішні службові записи не використовуються.

Таке представлення дає змогу впорядкувати вихідний матеріал та усунути неоднозначності, що можуть заважати подальшому аналізу. Після

обробки кожен діалог зберігався у вигляді набору структурованих записів у таблицях.

Формалізація вхідних даних у вигляді CSV із фіксованими полями дозволяє розглядати кожен діалог як структурований об’єкт, придатний для подальшої алгоритмічної обробки та повторного використання в експериментах.

Наступним етапом підготовки стало формування словників помилок. Для кожного з критеріїв — Dialogue Structure & Timings та Identification of the Customer’s Needs — були створені окремі таблиці бази даних, що містять:

- ідентифікатор запису у базі даних;
- назву помилки;
- розгорнутий опис;
- час створення запису.

Опис кожної помилки сформульований так, щоб бути достатнім для моделі й не містити двозначностей. Саме ці структуровані описи стали основою для побудови інструкцій до моделі та забезпечили чіткість експерименту.

Важливою частиною підготовки стала розробка інструкцій для роботи з мовною моделлю. Для кожного діалогу та кожного критерію створювався окремий запит, що містив:

- інструкцію для моделі щодо ролі (оцінювач якості);
- перелік помилок з відповідного словника;
- текст діалогу у компактному форматі;
- вимогу повернути структуру у вигляді суворого JSON.

Формат відповіді був зафіксований незалежно від змісту діалогу, що дозволяє моделі не відхилятися від заданої структури та полегшує подальший аналіз.

Підготовлений експериментальний процес має чітко визначену алгоритмічну структуру та реалізує послідовний автоматизований пайплайн

обробки даних. Алгоритм експериментального процесу було визначено як послідовність етапів:

1. Завантаження діалогу у CSV-форматі.
2. Парсинг і збереження у базу даних.
3. Формування текстового представлення діалогу.
4. Вибір критерію, зчитування відповідного словника помилок.
5. Формування запиту та передача його моделі gpt-4.1-mini.
6. Отримання JSON-відповіді та перевірка її структури.
7. Збереження результату у таблиці бази даних

Такий підхід дозволяє отримувати окремий незалежний результат для кожного критерію. Для досягнення відтворюваності всі параметри були зафіксовані:

- незмінний набір діалогів;
- стабільне серверне середовище;
- однакові параметри моделі gpt-4.1-mini;
- ідентичний формат інструкцій
- уніфікована структура збереження результатів.

Це дозволило гарантувати, що результати експерименту залежать саме від поведінки моделей та змісту діалогів, а не від випадкових чинників або зміни конфігурацій.

Таким чином, сукупність виконаних підготовчих дій забезпечила можливість коректного проведення експерименту та створила основу для аналізу результатів роботи MVP-системи.

4.3.Експериментальне середовище

В якості середовища для розгортання та виконання експерименту було обрано окремий віртуальний сервер, налаштований відповідно до вимог розробленої MVP-системи. Такий підхід забезпечує контрольованість умов експерименту, відтворюваність результатів та можливість ізоляції впливу зовнішніх чинників на роботу системи.

Серверне середовище було побудовано на операційній системі Ubuntu Server 22.04, що є стабільною платформою для вебзастосунків і дозволяє гнучко керувати конфігурацією. Для управління інфраструктурою було використано aaPanel (див.рис.4.1), що забезпечує зручний інтерфейс для роботи з вебсервером, базою даних та файловою структурою проєкту.

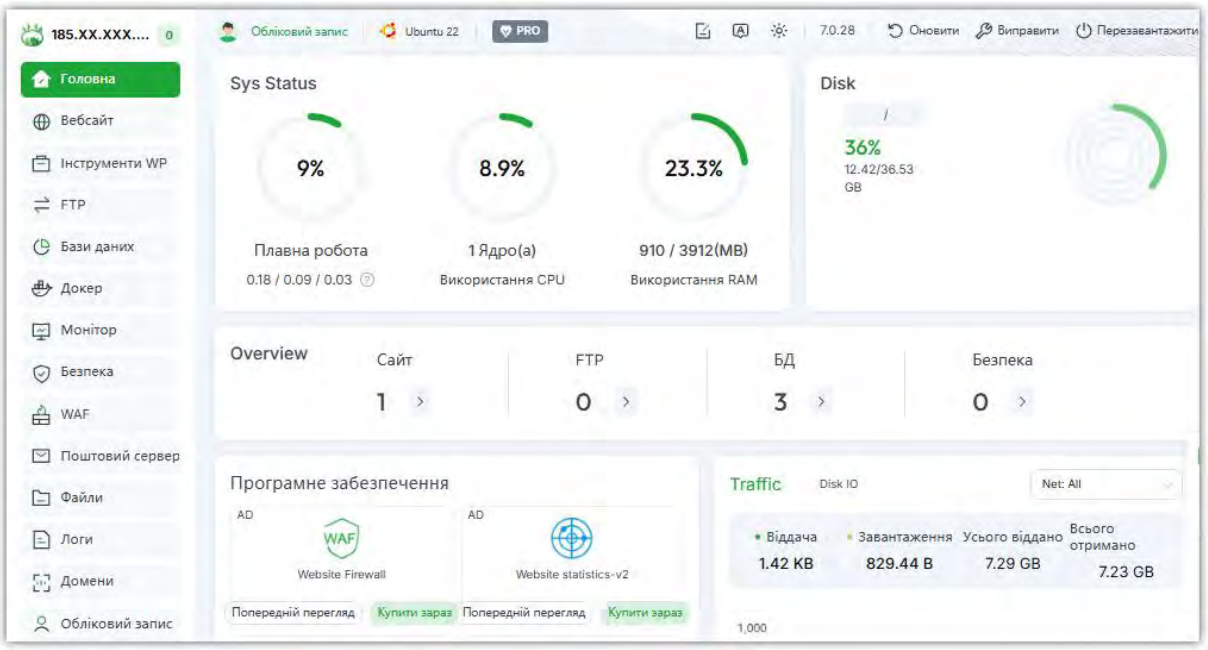

Рисунок 4.1 — головна панель aaPanel.

Для забезпечення коректної роботи MVP-системи було розгорнуто середовище із заздалегідь визначеним набором програмних компонентів. Їх перелік та призначення узагальнено у Таблиці 4.2.

Таблиця 4.2 — Компоненти середовища.

Компонент	**Версія**	**Призначення**
Ubuntu Server	22.04	базова ОС для хостингу системи
Nginx	1.24	вебсервер для прийому запитів
PHP	8.3	серверна логіка, обробка CSV, виклики LLM
MySQL	8.0.35	зберігання діалогів, помилок, результатів
aaPanel	7.0.28	керування сервером, файлами, доменом
phpMyAdmin	5.2	керування базою даних

Структура бази даних, що використовується в MVP, зображена на Рисунку 4.2. Вона включає таблиці діалогів, повідомлень та результатів аналізу, а також окремі словники помилок для кожного критерію оцінювання.

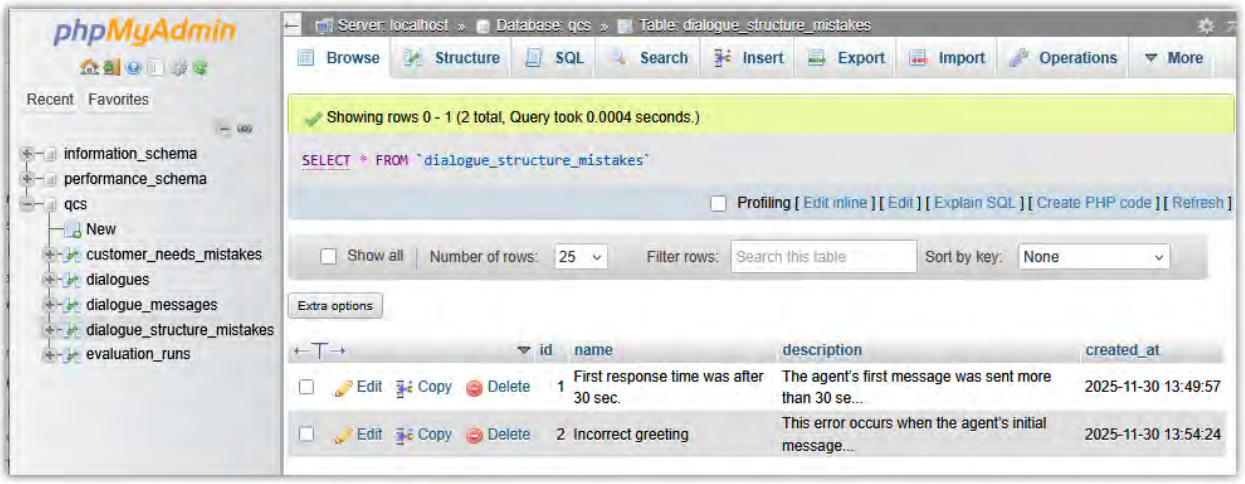

Рисунок 4.2 — phpMyAdmin зі структурою таблиць.

Кодова частина MVP-системи була розгорнута у каталозі вебзастосунку. Файлова структура побудована за принципом відокремлення конфігурацій, бібліотек, службових функцій та інтерфейсу. Це забезпечує модульність, відокремлення логіки обробки даних від взаємодії з користувачем та повну прозорість робочого процесу. На рисунку 4.3 наведено фактичну структуру директорій проєкту на сервері.

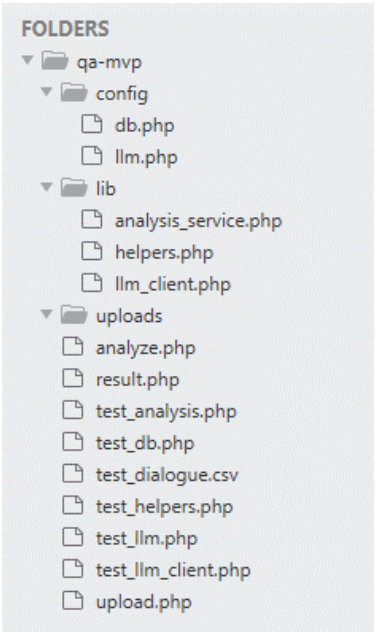

Рисунок 4.3 — Структура файлів MVP-системи (фрагмент із Sublime).

Структура файлів відображає розподіл функцій у системі: конфігурація середовища, модулі логіки аналізу, сервіс взаємодії з мовною моделлю, а також сценарії, що реалізують етапи експерименту (завантаження діалогу, виклик

аналізу, відображення результатів). У Таблиці 4.3 наведено стислий опис ролі кожного файлу.

Таблиця 4.3 – Основні файли MVP та їх призначення.

Файл	Призначення
config/db.php	Параметри підключення до бази даних MySQL;
config/llm.php	Налаштування взаємодії з мовною моделлю (endpoint, модель, ключ).
lib/llm_client.php	Єдиний клієнт для виклику LLM; формує та надсилає запити.
lib/helpers.php	Допоміжні функції: парсинг CSV, збереження діалогу, завантаження помилок тощо.
lib/analysis_service.php	Логіка аналізу діалогу: побудова промптів, отримання та збереження JSON результатів.
upload.php	Форма та логіка завантаження CSV у систему.
analyze.php	Запуск LLM-аналізу для обраного діалогу за всіма критеріями.
result.php	Відображення результатів аналізу у зручному для читання вигляді.
uploads/*.csv	Тестові діалоги, використані у рамках експерименту.

Узгодженість роботи компонентів середовища забезпечується чіткою роллю кожного модуля:

- MySQL зберігає структуровані дані;
- PHP отримує і формує текст діалогу, звертається до LLM, фіксує JSON-відповідь;
- Nginx забезпечує обробку запитів та взаємодію користувача з інтерфейсом;
- aaPanel полегшує адміністрування, логування та відстеження стабільності експериментального середовища.

У межах експериментального середовища передбачено мінімальний веб-інтерфейс для завантаження CSV-файлу з діалогом. Інтерфейс дозволяє:

- обрати файл
- переглянути його структуру перед збереженням
- ініціювати автоматичний аналіз LLM.

Описане експериментальне середовище забезпечує автоматизований цикл контролю якості комунікацій після ініціації аналізу, включно з обробкою діалогу, застосуванням формалізованих підкритеріїв, виконанням покрокового reasoning мовною моделлю, формуванням структурованих результатів у

форматі JSON та їх збереженням у базі даних. Завантаження діалогів у систему здійснюється вручну з метою забезпечення керованості та відтворюваності експерименту, однак усі подальші етапи оцінювання виконуються без участі людини. Такий підхід дозволяє чітко відокремити етап ініціації від автоматизованої логіки контролю якості та підтверджує, що розроблена MVP є саме автоматизованою системою контролю, а не аналітичним або рекомендаційним інструментом з ручним втручанням у процес прийняття рішень.

Для відтворюваності експерименту всі версії компонентів були зафіксовані, а конфігурація середовища повністю задокументована. Це дозволяє за необхідності повторити експеримент на іншому сервері з аналогічними параметрами.

Узгоджена структура середовища забезпечила стабільні умови для запуску MVP та дала можливість гарантувати, що результати експерименту визначаються логікою автоматизованої системи контролю якості, а не особливостями інфраструктури чи діями користувача.

4.4. Результати експерименту

Експериментальне дослідження було спрямоване на перевірку коректності роботи MVP-системи під час автоматизованого аналізу п’яти реалістичних діалогів між клієнтами та агентами служби підтримки.

Основними завданнями експерименту були:

1. перевірка стабільності формування структурованих відповідей мовної моделі у заданому форматі;
2. оцінка здатності системи автоматично виявляти порушення стандартів комунікації за критеріями Dialogue Structure & Timings та Identification of the Customer’s Needs;
3. зіставлення результатів автоматизованого аналізу з експертною інтерпретацією діалогів;

4. виявлення обмежень чинної конфігурації критеріїв та визначення напрямів їх подальшого розширення..

Для кожного діалогу система в автоматичному режимі виконувала аналіз за кожним критерієм окремо, формуючи JSON-висновок, який містив:

- ознаку наявності або відсутності помилки;
- релевантні цитати з діалогу;
- текстовий коментар моделі щодо прийнятого рішення.

На основі збережених JSON-результатів веб-інтерфейс відображав підсумки аналізу у табличному вигляді. Повні скриншоти результатів, журнали виконання та вихідні тексти діалогів наведено у Додатку С, що забезпечує відтворюваність експерименту та можливість незалежної перевірки отриманих висновків..

4.4.1. Аналіз діалогів за критерієм "Dialogue Structure & Timings"

У межах першого критерію автоматизовано оцінювалися такі аспекти:

- коректність початкового привітання агента (наявність вітання, представлення, звернення до клієнта та пропозиції допомоги);
- час першої відповіді агента відносно моменту першого повідомлення клієнта (граничне значення — 30 секунд).

Результати аналізу засвідчили:

1. У трьох із п'яти діалогів система коректно визначила відсутність порушень у структурі привітання агента.
2. У чотирьох діалогах модель правильно інтерпретувала часові інтервали між повідомленнями та сформувала адекватний пояснювальний коментар щодо фактичної затримки відповіді.
3. У двох діалогах (№2 та №4) зафіксовано повторювану аномалію: за відсутності порушення таймінгу модель формувала коректний текстовий коментар, проте одночасно встановлювала булеву ознаку «Так», що формально вказує на наявність помилки.

Такі випадки свідчать про недостатню жорсткість інструкції щодо синхронізації логічного висновку з табличною міткою. Модель коректно виконує семантичний аналіз, однак під час генерації структурованого результату може формувати булеві значення механічно. Подібна поведінка є типовою для LLM за відсутності явної вимоги логічної узгодженості між коментарем і формальною оцінкою.

Водночас у діалозі №5 система коректно виявила реальне порушення таймінгу, оскільки час відповіді агента становив 42 секунди, що перевищує встановлений норматив. Це підтверджує здатність системи адекватно працювати в однозначних сценаріях без суперечливого контексту.

4.4.2. Аналіз діалогів за критерієм "Identification of the Customer's Needs"

Другий критерій був спрямований на оцінку здатності агента:

- ставити релевантні уточнювальні запитання;
- уникати запитань, що не впливають на вирішення проблеми;
- демонструвати розуміння суті запиту клієнта.

За результатами експерименту встановлено:

1. У двох діалогах (№3 та №5) система коректно визначила, що всі запитання агента є релевантними, а зайві уточнення відсутні.
2. У діалогах №1 та №4 модель коректно виявила відсутність необхідних уточнювальних питань (наприклад, у випадку фінансової операції або збільшення кредитного платежу).
3. У діалозі №4 система правильно класифікувала "Unnecessary questions", оскільки агент ставив запитання, які не мали стосунку до суті проблеми, що повністю збіглося з експертною оцінкою.
4. Діалог №1 показав специфічний випадок: діалог був обірваним, і агент фізично не мав можливості поставити уточнювальні запитання. Модель класифікувала це як помилку, хоч з точки зору експертного оцінювання таке трактування є хибнопозитивним.

Цей результат ілюструє, що модель не завжди враховує контекст неповного діалогу і схильна інтерпретувати відсутність уточнень як недопрацювання агента, навіть коли це об’єктивно неможливо.

Показовим є діалог №3, у якому ШІ не зафіксувала жодної помилки, оскільки з формальної точки зору:

- привітання відповідає стандарту,
- таймінги коректні,
- уточнюючі питання релевантні.

Однак експертний аналіз виявив критично важливу проблему: агент ініціював фінансове рішення (повернення коштів) без додаткової верифікації особи клієнта та без підтвердження операції.

Таке рішення створює ризики безпеки, однак жоден із поточних критеріїв MVP не охоплює аспектів комплаєнсу та ідентифікації клієнта. Відповідно, навіть коректна робота моделі не дає змоги виявити порушення.

Цей кейс показує необхідність розширення бази критеріїв, зокрема додавання категорій, пов'язаних із:

- фінансовою безпекою,
- процедурою верифікації,
- дотриманням внутрішніх інструкцій служби підтримки.

У ході експерименту встановлено, що:

- у всіх п’яти випадках система формувала структуровано валідний JSON без синтаксичних помилок;
- формат відповідей був стабільним для обох критеріїв;
- виявлені суперечності між текстовими коментарями та булевими мітками мають системний характер і зумовлені особливостями інструкції;
- обробка CSV-файлів та збереження результатів у таблиці evaluation_runs виконувалися без збоїв;
- помилки API або інфраструктурні відмови під час експерименту не зафіксовані.

Отримані результати підтверджують технічну надійність MVP та демонструють ключову особливість розробленої системи: мовна модель не здійснює узагальнену суб'єктивну оцінку діалогу, а автоматизовано декомпонує його на формалізовані елементи відповідно до заданих критеріїв, що забезпечує прозорий, відтворюваний та масштабований контроль якості комунікацій.

Висновки до розділу 4

Проведений експеримент підтвердив працездатність MVP-системи автоматизованого контролю якості комунікацій, побудованої на багаторівневій моделі оцінювання, що включає критерії, підкритерії та формалізовані індикатори порушень, а також використанні мовної моделі. Усі діалоги були коректно оброблені, результати формувалися у стабільному JSON-форматі, а система здебільшого правильно визначала порушення у структурі привітання, таймінгах відповіді та якості уточнення потреб клієнтів.

Разом з тим виявлено кілька закономірностей. Модель продемонструвала здатність до точного семантичного аналізу, однак у двох випадках спостерігалася неузгодженість між поясненням і булевою позначкою — наслідок недосконалої інструкції, що потребує доопрацювання. Окремі сценарії показали, що обмежений набір критеріїв не дозволяє виявити важливі недоліки, зокрема пов'язані з фінансовою безпекою чи верифікацією клієнта, що було підтверджено шляхом порівняння автоматизованих результатів з експертною оцінкою. Також зафіксовані хибнопозитивні спрацювання в умовах обірваних або неповних діалогів.

Загалом експеримент підтвердив коректність обраної архітектури та її здатність виконувати підкритерійний аналіз, а не давати загальну суб'єктивну оцінку діалогу, що свідчить про виконання поставлених завдань дослідження. Це створює основу для подальшого розширення критеріїв, уточнення інструкцій і переходу до повноцінної автоматизованої системи контролю якості обслуговування.

5. РОЗРОБКА СТАРТАП ПРОЕКТУ «АВТОМАТИЗОВАНА СИСТЕМА КОНТРОЛЮ ЯКОСТІ КОМУНІКАЦІЇ МІЖ КЛІЄНТАМИ І ПЕРСОНАЛОМ У КОНТАКТ-ЦЕНТРУ»

5.1. Опис та технологічний аудит ідеї стартап-проєкту

Структура аналізу стартап-проєкту побудована відповідно до загальних рекомендацій щодо розроблення інноваційних проєктів [15]. Ідея стартап-проєкту ґрунтується на результатах попередніх теоретичних і практичних розділів, у яких було обґрунтовано потребу у сучасній системі автоматизованого контролю якості комунікацій між клієнтами та персоналом контакт-центру. Основою інноваційності є створення продукту, здатного забезпечувати повномасштабний аналіз діалогів із застосуванням великих мовних моделей з подальшим формуванням доказових фрагментів, зворотного зв'язку та інтеграції з корпоративними стандартами. Такий підхід дозволяє перетворити контроль якості з вибіркового та трудомісткого процесу на безперервну та технологічно обґрунтовану функцію управління сервісом.

Таблиця 5.1 – Опис ідеї стартап-проєкту.

Зміст ідеї	Напрямки застосування	Вигоди для користувача
Автоматизована система контролю якості комунікацій між клієнтами та персоналом контакт-центру, що використовує великі мовні моделі для глибинного підкритерійного аналізу діалогів та формування evidence-фрагментів як доказової основи оцінювання.	1. Масове оцінювання діалогів контакт-центрів (голос, чат).	100% охоплення діалогів.
	2. Контроль якості роботи агентів у режимі регулярного аудиту.	Прозорість та обґрунтованість оцінок.
	3. Дотримання внутрішніх стандартів сервісу та регуляторних вимог.	Скорочення витрат на ручний QA у 4–7 разів.
	4. Автоматизована аналітика при інтеграції з CRM-системами.	Прискорення покращення сервісу за рахунок точних рекомендацій для агентів.

Сучасні компанії стикаються з низкою проблем, пов'язаних із традиційним аудитом розмов: низьким охопленням (як правило, 10–15 % діалогів), високою трудомісткістю, суб'єктивністю оцінок і складністю повторюваного аналізу на великих обсягах даних. За даними галузевих досліджень, зростання ринку автоматизації контакт-центрів відбувається зі

швидкістю до 20 % на рік, що підтверджує актуальність рішень, які спираються на новітні підходи до обробки природної мови та впровадження штучного інтелекту [3]. У таких умовах стартап, спрямований на впровадження моделі глибинного аналізу комунікацій із використанням LLM, має високий потенціал як у технологічному, так і в комерційному вимірах.

Для визначення конкурентоспроможності ідеї стартап-проєкту проведено порівняння техніко-економічних характеристик із пропозиціями основних конкурентів. Аналіз у Таблиці 5.2 дозволяє окреслити сильні (S), нейтральні (N) та слабкі (W) сторони ідеї щодо існуючих рішень.

Таблиця 5.2 – Порівняльна характеристика техніко-економічних властивостей ідеї стартап-проєкту.

№	Техніко-економічні характеристики	Мій проєкт	MaestroQA	Scorebuddy	ZendeskQA	W	N	**S**
1	Можливість підкритерійного аналізу	є	частково	немає	частково			✓
2	Фрагменти доказових цитат	є	немає	немає	частково			✓
3	Інтеграція з CRM / власними моделями	є	є	частково	є		✓	
4	Локальне розгортання (преміум пакет)	є (квантовані моделі)	немає	немає	немає	✓		

Інформаційна карта стартап-проєкту

Перед початком технологічного аудиту доцільно представити цілісну інформаційну характеристику ідеї у вигляді узагальнюючої карти Таблиця 5.3, яка відображає ключові параметри проєкту та структурує його сутність.

Таблиця 5.3 – Інформаційна карта стартап-проєкту.

Параметр	Зміст
Призначення продукту	Автоматизований аналіз діалогів контакт-центру з виявленням порушень, формуванням рекомендацій і доказових фрагментів
Цільові користувачі	Супервайзери, аналітики якості, менеджери контакт-центрів, команди навчання персоналу

Продовження таблиці 5.3 – Інформаційна карта стартап-проєкту.

Параметр	Зміст
Технологічна база	Великі мовні моделі, транскрипція, семантичний аналіз, база знань, вебінтерфейс для аналітики
Унікальна цінність	Повне охоплення комунікацій, прозорість оцінок, зменшення витрат на ручний аудит, можливість адаптації під стандарти компанії
Потенційні ризики	Вартість мовних моделей, вимоги до конфіденційності, варіативність діалогів, інтеграційні обмеження

Представлена карта дозволяє систематизувати бачення продукту та відобразити його ключові властивості, що надалі використовується для проведення технологічного аудиту.

Морфологічна карта розвитку продукту

Морфологічний аналіз у Таблиці 5.4 застосовано для дослідження можливих траєкторій розвитку продукту та визначення придатних комбінацій технологічних і функціональних параметрів. Підхід дає змогу оцінити варіативність рішень і вибрати оптимальну конфігурацію MVP.

Таблиця 5.4 – Морфологічна карта розвитку продукту.

Параметр	Варіант 1	Варіант 2	Варіант 3	Варіант 4
Канал аналізу	Голосові дзвінки	Чат	E-mail	Омніканальність
Тип моделі	Хмарна LLM	Локальна модель	Комбінована схема	Квантизована модель
Логіка аналізу	Підкритерійний аналіз	Повний контекст	Аналіз за метриками	RAG-доступ до бази знань
Масштабованість	Вибірковий аналіз	50 % діалогів	100 % діалогів	Паралельна обробка
Формат взаємодії	Вебінтерфейс	API	Інтеграція з CRM	Внутрішня панель контролю

З аналізу таблиці випливає, що найбільш ефективною конфігурацією для MVP є використання хмарної мовної моделі з підкритерійним аналізом і доступом до корпоративної бази знань, оскільки це забезпечує баланс між точністю, швидкістю розгортання і відсутністю потреби в потужній локальній інфраструктурі. Надалі продукт може розвиватися в напрямі квантизованих

локальних моделей, що зменшують вартість обчислень і дають змогу компаніям контролювати конфіденційність даних.

Технологічний аудит ідеї стартапу

Технологічний аудит спрямований на оцінювання життєздатності ідеї з огляду на поточний стан ринку, конкурентні обмеження та технічні ресурси, необхідні для реалізації проєкту. У сучасних контакт-центрах застосовуються інструменти аналітики розмов, однак значна частина з них базується на фіксованих правилах або традиційних моделей машинного навчання. Такі системи недостатньо адаптивні, потребують великих обсягів даних для навчання та не забезпечують високої пояснюваності результатів, що є критичним для аудиту якості обслуговування.

Ідея стартапу має значну технологічну перспективу з огляду на те, що великі мовні моделі демонструють здатність до контекстної інтерпретації, формування логічних пояснень і проведення аналізу на підставі різнорідних інформаційних фрагментів. Окрім цього, архітектура рішення передбачає можливість поєднання семантичних обчислень з корпоративною базою знань, що дозволяє адаптувати систему до специфіки кожного контакт-центру.

Рівень конкуренції у сфері автоматизованого контролю якості є середнім. На міжнародному ринку представлені рішення, які пропонують автоматичну аналітику, проте більшість з них орієнтована на універсальні сценарії та не реалізує глибинний підкритерійний аналіз, який дозволяє виявляти структурні та комунікативні помилки в діалозі. Переважна частина систем не забезпечує прозорості й не надає доказових фрагментів, що знижує довіру користувачів до результатів автоматичного оцінювання. Наявність такої прогалини формує сприятливе середовище для запуску рішення, що пропонує пояснюваність, індивідуальне налаштування та інтеграцію з внутрішніми стандартами компанії.

Серед бар’єрів входу можна виокремити складність обробки конфіденційних даних, необхідність оптимізації витрат на використання мовних моделей і потребу в адаптації системи до різних галузей і форм

комунікації. У той самий час ці бар'єри стають одночасно конкурентними перевагами, оскільки команди, здатні їх подолати, отримують значно вищу ринкову позицію. Додатковою перевагою є те, що розробка MVP не вимагає великих інвестицій на етапі концепції та може бути реалізована в умовах обмежених ресурсів, оскільки використовує хмарну інфраструктуру та мінімально необхідний набір модулів для перевірки гіпотез.

Таблиця 5.5 — Технологічна здійсненність ідеї стартап-проєкту.

№	Ідея проєкту	Технології її реалізації	Наявність технологій	Доступність технологій авторам
1	Автоматизоване оцінювання діалогів	LLM-моделі (GPT-клас, Llama, Claude)	Технології існують і доступні через API	Доступні (OpenAI API, локальні моделі)
2	Підкритерійний аналіз та evidence-фрагменти	Техніки NLU, chain-of-thought, prompting	Технології існують, адаптація мінімальна	Доступні розробникам
3	Масштабована обробка великих обсягів даних	Хмарні платформи (AWS, Azure, GCP), serverless-архітектура	Існують	Доступні за моделлю pay-as-you-go
4	Інтеграція з CRM та контакт-центрами	REST API, вебхуки, SDK CRM-платформ	Існують	Доступні (Freshdesk, Creatio, Zendesk)
5	Локальне розгортання за вимогою клієнта	Квантовані моделі, контейнеризація (Docker)	Частково існують, потребують доопрацювання	Частково доступні, розвиток можливий
6	Зберігання та аналітика результатів	PostgreSQL / ClickHouse / OLAP-схеми	Існують	Доступні авторам

Розроблення MVP стартап-проєкту

Мінімально життєздатний продукт створюється з метою перевірки технічних і ринкових гіпотез стартапу, а також демонстрації здатності системи здійснювати автоматизований аналіз комунікацій у реальних умовах. MVP охоплює базові модулі та алгоритми, які забезпечують цінність для користувача і дозволяють перевірити ключові припущення щодо точності, стабільності й практичної корисності рішення.

Програмна архітектура MVP побудована на модульному принципі. До її складу входять інтерфейс завантаження діалогів, модуль підготовки та структурування даних, блок обробки мовною моделлю, база даних для зберігання результатів і модуль візуалізації оцінок. Такий підхід забезпечує розширюваність і можливість безболісного додавання нових критеріїв та індикаторів аналізу. Архітектура також містить механізм порівняння дій агента з еталонними стандартами через використання бази знань, що дає змогу проводити індивідуальні перевірки відповідно до політик конкретної компанії.

Основною складовою MVP є алгоритм підкритерійного аналізу діалогів, який дозволяє розділяти загальну оцінку на низку структурних і поведінкових параметрів. Мовна модель отримує діалог, критерій і короткий план оцінювання, після чого аналізує відповідність, формує пояснення та виділяє доказові фрагменти. Такий підхід має низку переваг, зокрема підвищену стабільність результатів, можливість відтворення логіки оцінювання та використання різних моделей без втрати консистентності.

На рівні практичної реалізації MVP включає зручний інтерфейс, що дозволяє завантажувати файли, переглядати розмови й оцінки, а також отримувати зведений звіт. Система функціонує в автоматичному режимі та не потребує участі експерта на етапі аналізу, що значно скорочує трудові витрати. У подальшому MVP може бути інтегрований із зовнішніми системами керування взаєминами з клієнтами, інструментами для коучингу персоналу та платформами управління контакт-центром.

З позиції технічної ефективності MVP демонструє здатність масштабуватися залежно від обсягів діалогів, підтримує паралельну обробку, а також може бути адаптований під вимоги безпеки, включно з локальним розгортанням. У міру розвитку проєкту можливе впровадження квантизованих моделей для зменшення вартості обчислень, а також удосконалення алгоритмів валідації результатів із застосуванням статистичних методів, таких як кореляційний аналіз або перевірка міжоцінювальної узгодженості.

Розроблений MVP підтверджує життєздатність ідеї стартапу та створює основу для подальшої комерціалізації, оскільки доводить можливість автоматизованого й відтворюваного аналізу діалогів зі збереженням точності, пояснюваності та адаптивності до потреб контакт-центрів.

5.2. Аналіз ринкових можливостей запуску стартап-проєкту

Оцінювання ринкових можливостей стартапу, спрямованого на автоматизацію контролю якості комунікацій у контакт-центрах, передбачає аналіз попиту, характеристик ринку, динаміки його розвитку та можливих обмежень, що можуть впливати на комерційну реалізацію продукту. Галузь контакт-центрів упродовж останніх років демонструє стале зростання обсягів звернень і поступове зміщення компаній у бік цифровізації процесів обслуговування. Зростання кількості клієнтських комунікацій породжує потребу в масштабованих інструментах контролю якості, що не залежать від людського фактора та забезпечують повне охоплення розмов. Водночас підвищення стандартів сервісу робить ручний аудит неефективним, що відкриває значні можливості для автоматизованих рішень на основі мовних моделей.

Наведені тенденції дають підстави припустити, що ринок є перспективним для впровадження інноваційного продукту, здатного вирішувати як проблему якості, так і проблему продуктивності операційного персоналу. Зростання інтересу компаній до технологій штучного інтелекту також сприяє позитивному сприйняттю інструментів цього класу, що усуває бар’єри первинної довіри та значно пришвидшує процес ринкового прийняття.

Попередня характеристика потенційного ринку

Характеристика ринку наведена у таблиці 5.6, що відображає ключові показники його стану та дозволяє оцінити ступінь привабливості для входу.

Таблиця 5.6 – Попередня характеристика ринку автоматизованих систем контролю якості комунікацій.

№	Показник	Характеристика
1	Кількість головних гравців	Середня; ринок представлений кількома міжнародними платформами та локальними рішеннями
2	Загальний обсяг продажів	Сегмент аналітики контакт-центрів демонструє приріст понад 15–20 % щороку
3	Динаміка розвитку	Зростає; впровадження систем AI-аудиту стає стандартом у великих компаніях
4	Наявність бар'єрів входу	Технологічні та регуляторні; потреба в обробці конфіденційних даних
5	Вимоги до сертифікації	Високі вимоги щодо безпеки та роботи з персональними даними
6	Середня норма рентабельності	Достатньо висока для SaaS-продуктів у сегменті B2B

Аналіз показує, що ринок перебуває на етапі зростання, а вхідні бар'єри не є надмірними для невеликих команд, здатних поставити продукт зі значною технологічною перевагою. Висока рентабельність і зростання попиту свідчать про доцільність подальшої комерціалізації.

Характеристика потенційних клієнтів та їх вимог

Потреба у повному та автоматичному аудиті формує широку базу споживачів, що включає як невеликі контакт-центри, так і великі аутсорсингові компанії та підприємства з масовими клієнтськими зверненнями. Узагальнена характеристика ключових сегментів наведена у таблиці 5.7.

Таблиця 5.7 – Характеристика потенційних клієнтів стартап-проєкту.

№	Потреба ринку	Цільова аудиторія	Відмінності у поведінці	Основні вимоги
1	Контроль якості обслуговування	Контакт-центри 50–500 операторів	Орієнтація на зниження витрат та стандартизацію процесів	Точність оцінювання, прозорість логіки
2	Підвищення продуктивності QA-команд	Аутсорсингові компанії	Високі вимоги до масштабованості	Автоматизація рутинних перевірок
3	Дотримання стандартів та нормативів	Банківські та фінансові установи	Підвищений рівень регуляторного контролю	Підтримка бази знань, аудит відповідності
4	Поліпшення досвіду клієнта	E-commerce, логістика	Важливість аналізу негативних звернень	Повне охоплення комунікацій, інтерпретованість

Споживачі у всіх сегментах очікують прозорих пояснень оцінок, гнучких можливостей налаштування та прогнозованої вартості продукту.

Фактори загроз для впровадження стартапу

Надалі виконано групування факторів, які можуть перешкоджати реалізації проєкту. Згідно з методичними вимогами, фактори подано у таблиці 5.8 разом із можливою реакцією компанії.

Таблиця 5.8 – Фактори загроз.

№	Фактор	Зміст загрози	Можлива реакція компанії
1	Вартість використання LLM	Значні витрати при великих обсягах даних	Оптимізація запитів, підтримка локальних моделей
2	Конкуренція з міжнародними платформами	Наявність сильних гравців на глобальному ринку	Диференціація через explainability та адаптивність
3	Обмеження щодо конфіденційності	Вимоги щодо зберігання та обробки даних	Локальні розгортання, політики безпеки
4	Нерівномірність якості транскрипції	Мовні та акустичні особливості	Використання декількох моделей розпізнавання
5	Складність інтеграції з CRM	Неоднорідність інфраструктури клієнтів	Розроблення універсального API-шару

Загрози не є критичними і можуть бути пом’якшені завдяки гнучкості технологій, що підтверджує реалістичність виходу на ринок.

Фактори можливостей

Ринок створює низку можливостей, що можуть бути використані для прискорення зростання стартапу, узагальнено покажимо у вигляді таблиці 5.9.

Таблиця 5.9 – Фактори можливостей.

№	Фактор	Зміст можливості	Можлива реакція компанії
1	Швидке зростання ринку AI-сервісів	Компанії активно шукають рішення для автоматизації	Позиціонування як спеціалізованого інструменту QA
2	Потреба у скороченні витрат	Компанії прагнуть зменшити навантаження на QA-відділи	Пропозиція повного охоплення діалогів
3	Перехід до сервісних моделей	Інтерес до SaaS-рішень із прогнозованими витратами	Гнучка система підписки
4	Підвищення ролі даних у прийнятті рішень	Бізнес потребує структурованих аналітичних даних	Інтегровані звіти та показники
5	Поширення омніканальних контакт-центрів	Зростає варіативність каналів комунікації	Підтримка голосу, чату та текстових каналів

Можливості свідчать про широкий простір для росту, якщо продукт адаптується до потреб різних сегментів.

SWOT-аналіз стартап-проєкту

На основі проведеного ринкового аналізу в таблиці 5.10 узагальнено сильні та слабкі сторони, можливості й загрози для реалізації проєкту.

Таблиця 5.10 – SWOT-аналіз стартапу.

Сильні сторони	Слабкі сторони
Висока точність аналізу, пояснюваність результатів, можливість адаптації під стандарти	Вартість LLM-обчислень, потреба в інтеграції з інфраструктурою клієнта
Можливості	Загрози
Зростаючий ринок, попит на автоматизацію, дефіцит якісних QA-інструментів	Конкурентні рішення, регуляторні обмеження, ризики роботи з даними

SWOT-матриця підтверджує, що сильні сторони продукту добре корелюють із потребами ринку. Основні слабкі сторони пов’язані з технологічними витратами, які можуть бути оптимізовані. Загрози не створюють критичних обмежень і компенсуються суттєвими ринковими можливостями.

5.3. Розроблення ринкової стратегії та маркетингової програми проєкту

Розроблення ринкової стратегії проєкту

Розробка ефективної ринкової стратегії розпочинається з чіткого визначення стратегії охоплення ринку. Першим кроком є вибір цільових груп потенційних споживачів, які формують попит на інноваційний продукт та визначають вимоги до його функціональності й цінності. Для стартап-проєкту, присвяченого автоматизації контролю якості комунікацій, основними сегментами виступають компанії з високим обсягом клієнтських звернень та потребою у стандартизованому аналізі діалогів. Характеристика цільових груп подана у таблиці 5.11.

Таблиця 5.11 – Характеристика цільових груп потенційних споживачів.

Опис групи	Готовність сприйняти продукт	Попит у сегменті	Конкурентність	Простота входу
Контакт-центри 50–300 агентів	Висока; активний перехід до автоматизації	Значний	Середня	Висока
Аутсорсингові контакт-центри	Висока; потреба у масштабуванні без витрат	Дуже високий	Висока	Середня
Банківські та страхові компанії	Висока; регуляторні вимоги до якості	Стабільний	Висока	Низька
E-commerce та логістика	Середня; акцент на швидкість сервісу	Зростаючий	Середня	Висока

Аналіз показує, що найбільш перспективними сегментами є середні та великі контакт-центри, а також аутсорсингові оператори, які прагнуть зменшити навантаження на відділи контролю якості. Ці групи демонструють високу готовність сприйняти продукт, що обґрунтовує доцільність диференційованої стратегії охоплення ринку, коли пропозиція адаптується під окремі сегменти.

Подальшим етапом є визначення базової стратегії розвитку, що задає загальний напрям позиціонування стартапу в конкурентному середовищі. В умовах ринку, де конкурують як традиційні rule-based системи, так і окремі AI-рішення, найбільш ефективною є стратегія диференціації, яка передбачає створення унікальної цінності для клієнтів за рахунок відмінних властивостей продукту. Відповідну систему рішень подано в таблиці 5.12.

Таблиця 5.12 – Базова стратегія розвитку проєкту.

Альтернатива розвитку	Стратегія охоплення	Ключові конкурентні позиції	Базова стратегія
Вихід на сегменти середніх і великих контакт-центрів	Диференційована	Пояснюваність результатів, повне охоплення діалогів, персоналізація критеріїв	Стратегія диференціації

У межах цієї стратегії стартап орієнтується не на цінову конкуренцію, а на підвищену якість продукту та його здатність автоматизувати складні процеси аналізу, що традиційно вимагали значної кількості робочого часу експертів. Такий підхід відповідає класичним моделям стратегічного

управління, де диференціація визначається як спосіб створення унікальної позиції продукту за рахунок інноваційних характеристик [4].

Наступним елементом є стратегія конкурентної поведінки. В умовах ринку, де вже існують міжнародні рішення, але локальний сектор автоматизованих систем контролю якості все ще недостатньо розвинений, найбільш раціональною стає флангова стратегія. Вона полягає у фокусуванні на тих аспектах, які конкуренти покривають частково або недостатньо — підвищена прозорість результатів оцінювання, адаптація під локальні стандарти, робота з корпоративними політиками. Така стратегія дозволяє уникнути прямої конкуренції з великими міжнародними системами та створити власну стійку нішу.

Таблиця 5.13 – Базова стратегія конкурентної поведінки стартап-проєкту.

Характеристика	Зміст
Першопрохідництво	Часткове: продукт впроваджує підхід, який ще не представлений серед локальних рішень, але існує на глобальному рівні
Характер поведінки на ринку	Уникнення прямої конкуренції з глобальними платформами; концентрація на слабко покритих потребах локальних компаній
Чи копіює характеристики конкурентів	Ні; продукт формує унікальні властивості (пояснюваність, підкритерійний аналіз, локалізація), які не реалізовані у більшості конкурентів
Стратегія конкурентної поведінки	Флангова стратегія, що передбачає фокус на сегментах і функціях, де конкуренти є найменш сильними
Ключове обґрунтування	Недостатня пропозиція рішень з підвищеною деталізацією та прозорістю оцінювання, локальною адаптацією та прозорою логікою оцінювання; можливість сформувати стійку нішу

Наступним етапом формування ринкової стратегії є визначення стратегії позиціонування, що продемонстровано у Таблиці 5.14. Вона орієнтована на створення чіткої системи асоціацій, за якими продукт ідентифікуватиметься на ринку.

Таблиця 5.14 – Стратегія позиціонування стартап-проєкту.

Вимоги аудиторії	Базова стратегія	Ключові конкурентні позиції	Асоціації позиціонування
Точність, доказовість, масштабованість	Диференціація	Прозорий автоматичний аудит; індивідуальні стандарти оцінювання; повне охоплення діалогів	«Прозорий аудит», «Інтелектуальна аналітика сервісу»

Продовження таблиці 5.14 – Стратегія позиціонування стартап-проєкту.

Вимоги аудиторії	Базова стратегія	Ключові конкурентні позиції	Асоціації позиціонування
Інтеграція в CRM та операційний стек	Диференціація	Підтримка API-інтеграцій, передача оцінок та фрагментів у CRM, можливість включення результатів у workflow	«Безшовна інтеграція», «Аналітика в один клік»
Гнучкість у виборі технологічної бази	Диференціація	Можливість підключення власних моделей компанії (LLM, локальні моделі, mini-LLM), режим on-premise	«Контроль над даними», «Безпека та автономність»

У межах формування стратегії позиціонування враховано ключові потреби клієнтів, серед яких усе більше компаній віддають перевагу платформам, що не лише виконують автоматичний аудит комунікацій, а й можуть бути інтегровані в існуючі бізнес-процеси. Саме тому одним із стратегічних напрямів позиціонування визначено безшовну інтеграцію з CRM-системами, що дозволяє автоматично передавати результати оцінювання, виявлені фрагменти діалогів, сигнали ризику або рекомендації для агентів у внутрішні операційні інструменти компанії.

Другим важливим елементом позиціонування є можливість підключення власних мовних моделей компанії, включно з локально розгорнутими й оптимізованими моделями. Це відповідає тенденції до підвищення вимог щодо безпеки даних і створює унікальну конкурентну перевагу, оскільки більшість існуючих рішень не дозволяють клієнтам повністю контролювати інтелектуальну обробку даних.

Розроблення маркетингової програми стартап-проєкту

Маркетингова програма охоплює визначення ключових характеристик продукту, принципів ціноутворення, організації збуту та системи комунікацій. У таблиці 5.15 показаний перший крок формування концепції товару, що базується на виявлених потребах споживачів.

Таблиця 5.15 – Ключові переваги концепції потенційного товару.

Потреба	Вигода для споживача	Ключові переваги (з урахуванням відмінностей від конкурентів)
Автоматизація аудиту комунікацій	Скорочення часу перевірки та зменшення навантаження на QA	Поглиблена та структурована оцінка діалог, а не лише загальна оцінка
Дотримання внутрішніх стандартів обслуговування	Єдина логіка оцінювання для всіх агентів	Висока обґрунтованість та прозорість результатів, що підвищує довіру до результатів аудиту
Підвищення якості сервісу	Прозорі персоналізовані рекомендації агентам	Пояснюваність (Explainability), адаптація під стандарти клієнта
Масштабованість та стабільність процесів	Можливість аналізу 100% дзвінків і чатів	Повне охоплення розмов і можливість інтеграції з CRM / власними моделями

На відміну від конкурентів, продукт пропонує підвищену прозорість та обґрунтованість результатів оцінювання, що підсилює довіру користувачів і забезпечує вищу якість управлінських рішень.

Подальшим кроком є опис трьох рівнів моделі товару, що узагальнено в таблиці 5.16.

Таблиця 5.16 – Три рівні моделі товару.

Рівень	Сутність
Товар за задумом	Забезпечення високої якості комунікацій через їх автоматизований аналіз
Товар у реальному виконанні	Вебінтерфейс із модулями завантаження, аналізу, звітності; база знань; інтеграції
Товар із підкріпленням	Підтримка, оновлення, адаптація критеріїв, можливість локального розгортання

Окремою конкурентною перевагою є підтримка інтеграцій із CRM-системами та можливість підключення власних мовних моделей замовника, що забезпечує контроль над даними та відповідність корпоративним вимогам безпеки.

Далі визначаються межі ціни, що враховують ринок аналогів, замінників і дохідність клієнтів. Для SaaS-рішень типовою є модель щомісячної підписки з урахуванням кількості діалогів.

Таблиця 5.17 – Межі встановлення ціни.

Ціни на аналоги	Ціни на замінники	Дохідність цільового сегменту	Межі ціни (для стартап-продукту)
MaestroQA, Scorebuddy: 50–90$ за агента/місяць; Zendesk QA (Klaus): від 60$ за агента + плата за AI	Загальні LLM-інструменти (ChatGPT API, Claude API) мають низьку вартість, але не є повноцінними QA-системами	Великі компанії з високим операційним бюджетом; преміум сегмент з можливістю інвестувати в AI-автоматизацію	1) Модель “за діалоги”: 0.01–0.05$ за діалог (масштаб залежно від обсягу); 2) Модель “за агента”: 40–70$ за агента/місяць; Enterprise — індивідуальні тарифи

Визначення меж встановлення ціни для стартап-проєкту здійснюється з урахуванням вартості аналогічних рішень на ринку, наявних товарів-замінників та доходності цільового сегменту. Аналіз цінових моделей провідних міжнародних платформ у сфері автоматизованого контролю якості комунікацій свідчить, що вартість AI-модулів у таких системах, як MaestroQA, Scorebuddy чи Zendesk QA, коливається від 50 до 90 доларів США за одного агента на місяць, при цьому окремо враховується плата за використання мовних моделей.

На відміну від конкурентів, цінова модель не базується виключно на оплаті за доступ до AI, а на комплексній вартості послуги, що включає повний цикл автоматизованого контролю якості та інтеграцію з корпоративною інфраструктурою.

Для запропонованого стартапу обрано гібридну модель ціноутворення, що включає два взаємодоповнювальні режими. Перший — це модель “оплати за обсяг діалогів”, у межах якої компанія сплачує за кількість оброблених комунікацій. Такий підхід є економічно доцільним для великих організацій, де аналіз здійснюється на рівні потоків, а не окремих агентів. Друга складова — тариф “за агента”, який є релевантним для компаній, що використовують індивідуальне оцінювання та прив’язують якість обслуговування до конкретних співробітників.

Відповідно до цього підходу, нижня та верхня межі ціни формуються виходячи з преміального позиціювання продукту та високої

платоспроможності сегменту великих компаній. Для моделі “за діалоги” орієнтовна ціна становить від 0.01 до 0.05 долара США за один діалог, залежно від масштабу використання та обсягу пакетів. Для моделі “за агента” доцільним є встановлення діапазону від 20 до 50 доларів США за агента на місяць, що є нижчим або порівнянним з вартістю глобальних аналогів, але пропонує істотно вищу функціональність у частині доказового аналізу та адаптації під внутрішні стандарти компанії. Для корпоративних клієнтів із високими вимогами до безпеки передбачаються enterprise-тарифи, які включають можливість локальної інтеграції та підключення власних моделей.

Запропоновані межі забезпечують достатню гнучкість і дозволяють адаптувати цінову політику під різні сегменти ринку, водночас зберігаючи преміальність продукту та його орієнтацію на великі компанії, які прагнуть повної автоматизації контролю якості комунікацій.

Система збуту має забезпечувати контрольованість взаємодії з клієнтами та підтримувати високий рівень сервісу, що є необхідним для збереження унікальної цінності продукту та його преміального позиціонування.

Таблиця 5.18 – Формування системи збуту стартап-проєкту.

Поведінка клієнтів	Функції постачальника	Глибина каналу збуту	Оптимальна система збуту
Прагнення до швидкого впровадження та прозорої демонстрації цінності продукту	Проведення демо, презентацій, персоналізованих сесій для enterprise-клієнтів	Короткий канал (прямі продажі)	Прямі продажі через власний sales-відділ; участь у конференціях у сферах сервісу та entertainment
Потреба в інтеграції з CRM та контакт-центрами, слабка пропозиція аналітичних інструментів у партнерських платформах	Надання API, технічної підтримки інтеграцій, супровід партнерів	Середня глибина (через партнерів)	Співпраця з платформами Freshdesk, Creatio, телефоніями та CRM-провайдерами
Схильність enterprise-клієнтів до стабільності та безперервності сервісу	Забезпечення SLA, підтримка та постійне оновлення	Короткий канал (SaaS-модель)	100% SaaS (хмарне розгортання), централізоване керування оновленнями

Завершальним елементом є концепція маркетингових комунікацій, що формує принципи взаємодії із цільовими аудиторіями.

Таблиця 5.19 – Концепція маркетингових комунікацій.

Поведінка клієнтів	Канали комунікацій	Ключові позиції (меседжі)	Завдання маркетингових комунікацій	Концепція звернення
Потребують доказів ефективності рішення (демо, практичні докази)	Галузеві конференції у сферах entertainment і customer experience; професійні івенти (Customer Experience Day, CCaaS events)	100% поглиблене охоплення діалогів; прозорий доказовий аналіз	Демонстрація реальної ефективності продукту у порівнянні з ручним QA та конкурентами	«AI-аудит, який доводить свою цінність на практиці»
Орієнтовані на швидке впровадження та мінімальні інтеграційні витрати	Партнерські канали (Freshdesk, Creatio marketplace)	Підтримка CRM та можливість підключення власних моделей	Показати простоту інтеграції та швидкий старт у команді клієнта	«Інтеграція в один клік. Аналітика без затримок»
Цінують прозорість і хочуть бачити явні переваги над конкурентами	Product demo hub, Proof-of-Concept / free pilot, ROI-калькулятор, порівняльні таблиці	Підтримка власних моделей; гнучкість і контроль над даними	Надати інструменти для самостійної оцінки цінності продукту	«Зробіть свій QA-процес прозорим і керованим»

Стратегія маркетингових комунікацій спрямована на формування довіри до продукту та демонстрацію його переваг у преміальному B2B-сегменті. Враховуючи поведінку клієнтів, ключовими елементами програми стають демонстрації, пілотні впровадження, аналітичні матеріали та порівняльні таблиці, що дозволяють підприємствам самостійно оцінити ефективність AI-аудиту. Особливе значення мають партнерські канали, які забезпечують доступ до цільових аудиторій через екосистеми CRM, та участь у спеціалізованих подіях, де замовники шукають інноваційні інструменти для підвищення якості обслуговування.

5.4. Бізнес-модель реалізації стартап-проекту та оцінювання його економічної ефективності

Для узагальнення ключових елементів функціонування стартап-проєкту на рисунку 5.1 було сформовано бізнес-модель у форматі Canvas [16], що

дозволяє компактно відобразити логіку створення цінності, структуру ресурсів та механізми отримання доходів.

<table>
<tr><th>Ключові партнери</th><th>Ключові види діяльності</th><th>Ціннісна пропозиція</th><th>Взаємовідносини з клієнтами</th><th>Сегменти користувачів</th></tr>
<tr><td rowspan="3">Постачальники LLM-моделей (OpenAI, Anthropic, локальні моделі);
Провайдери хмарної інфраструктури;
CRM-платформи (Freshdesk, Creatio, Zendesk);
Партнери-резелери.</td><td>Розроблення NLP/LLM-модулів;
Підкритерійний аналіз та формування evidence-фрагментів;
Інтеграції з CRM і контакт-центрами;
Коучинговий зворотний зв’язок.</td><td rowspan="3">Наш продукт забезпечує 100% автоматизований аналіз діалогів у контакт-центрах із прозорим аудитом на основі доказових фрагментів. Рішення дозволяє скоротити витрати на QA у 4–7 разів, підвищити точність оцінювання та адаптувати систему під стандарти конкретної компанії.

Продукт створює цінність завдяки: обґрунтованим та пояснюваним результатам оцінювання; поглибленій деталізації діалогів на рівні окремих критеріїв; можливості інтеграції з CRM та існуючими процесами без змін у інфраструктурі; індивідуальному налаштуванню під правила й політики клієнта.</td><td>Підтримка B2B-клієнтів;
Персональний супровід при впровадженні;
24/7 технічна підтримка.</td><td rowspan="3">Enterprise контакт-центри;
Середні та великі сервісні компанії;
Банківський, e-commerce, телеком-сектори;
Інтертеймент-направлення.</td></tr>
<tr><td>Ключові ресурси</td><td>Канали збуту</td></tr>
<tr><td>Мовні моделі, база знань, серверна інфраструктура;
Інженери NLP/ML;
API та CRM-інтеграції;
База клієнтських діалогів.</td><td>Прямі продажі (sales);
Демонстрації, пілоти;
Партнери CRM-маркетплейсів.</td></tr>
<tr><td colspan="3">Структура витрат</td><td colspan="2">Джерела доходів</td></tr>
<tr><td colspan="3">Розробка LLM-модуля для інтеграції мовних моделей;
Витрати на інтерфейс адміністратора (користувача);
Хмарні ресурси, підтримка;
Продажі та маркетинг.</td><td colspan="2">Підписка SaaS;
Плата за інтеграції;
Оплата за обсяг аналізу;
Оплата за кількість агентів;
Enterprise-ліцензії.</td></tr>
</table>

Рисунок 5.1 – Бізнес-модель Canvas.

Реалізація стартап-проєкту передбачає формування команди, здатної забезпечити розроблення, впровадження та подальший розвиток інтелектуальної системи автоматизованого контролю якості комунікацій. Команда включає ключові компетенції у сфері управління продуктом, програмної інженерії, машинного навчання, аудиту комунікацій та продажів. Її структура є оптимальною для раннього етапу запуску продукту і забезпечує баланс між технічними можливостями та необхідністю демонстрації цінності рішення потенційним клієнтам.

Таблиця 5.20 – Команда стартап-проєкту.

Посада	Основні функції	Роль у проєкті
CEO / Product Owner	Стратегія, бачення продукту, інвестиції, управління пріоритетами	Керування розвитком проєкту та взаємодія з партнерами
CTO	Архітектура, бекенд, інфраструктура, безпека	Технічне керівництво та інтеграція LLM-складових
ML Engineer	Розроблення та налаштування моделей, якість аналізу	Забезпечення точності та стабільності AI-компонентів
QA/Business Analyst	Формування критеріїв оцінювання, валідація якості	Відповідність моделі стандартам клієнтів
Sales Engineer (пре-сейлз)	Проведення демо, презентацій, підготовка PoC	Забезпечення залучення клієнтів та конверсії лідів

Реалізація стартапу здійснюється відповідно до календарного плану, що охоплює дворівневу структуру робіт: технічну розробку та ринкову підготовку. Загальна тривалість циклу становить 12 місяців і включає дослідження ринку, створення MVP, тестування з реальними діалогами, інтеграцію з CRM, пілотне впровадження та масштабування. План побудований на основі контрольного списку завдань проєкту та узгоджений зі стратегією виходу на преміум-сегмент.

Таблиця 5.21 – Календарний план-графік реалізації стартапу.

Етап	Завдання	Тривалість	Ключовий результат
Дослідження та проєктування	Аналіз ринку, постановка вимог, архітектура MVP	1–2 міс.	Готова технічна та продуктова концепція
Розробка MVP	Розроблення модулів аналізу, UI, інтеграція LLM	3–4 міс.	Робочий MVP з основними функціями
Тестування та валідація	Перевірка якості, порівняння з експертами, виправлення	2 міс.	Стабільність роботи та точність аналізу
Інтеграції та PoC	CRM-підключення, тестові впровадження	2 міс.	Доведення цінності продукту
Маркетинг та продажі	Демонстрації, конференції, контент	2 міс.	Формування перших клієнтів

Організація “виробництва” цифрового продукту ґрунтується на SaaS-моделі, що забезпечує масштабованість і доступність для клієнтів незалежно від їхньої інфраструктури. Операційний цикл складається з обробки діалогів, аналізу мовною моделлю, формування оцінок за критеріями, створення доказових фрагментів, адаптації результатів до CRM клієнта та збереження їх

у внутрішній аналітичній системі. Така модель мінімізує операційні витрати та забезпечує швидке розгортання сервісу в enterprise-сегменті.

Бізнес-модель проєкту базується на поєднанні ціннісної пропозиції (автоматизований, прозорий та масштабований контроль якості), чітко визначених клієнтських сегментів (контакт-центри з високим навантаженням, аутсорсингові оператори, банки та страхові компанії), партнерських каналів доступу (Freshdesk, Creatio), а також гнучкої моделі монетизації. Джерелами доходів є два тарифи: оплата за обсяг діалогів у разі масового аналізу та оплата за агента, коли компанія здійснює детальний контроль ефективності окремих працівників. Для корпоративних клієнтів передбачено enterprise-пакети з розширеними функціями та можливістю підключення власних моделей.

Економічна частина передбачає аналіз витрат, необхідних для запуску стартапу, та визначення прогнозованих фінансових результатів. Загальна потреба в інвестиціях для створення MVP, хмарної інфраструктури, маркетингу та операційних витрат становить близько 100 000 доларів США, що є типовим показником для AI-стартапів із високою технологічною складовою. Прогнозований річний дохід стартапу за умов залучення 8–12 клієнтів у преміум-сегменті становить 300 000–500 000 доларів США, що дає можливість досягти точки беззбитковості протягом першого року роботи та забезпечити подальше масштабування продукту. Урахування моделі повторюваного доходу (SaaS) забезпечує стабільність фінансових результатів і передбачуваний розвиток компанії.

Таблиця 5.22 – Основні витрати та очікувані фінансові результати.

Стаття	Сума	Період	Коментар
Розробка MVP	$35 000–50 000	Разово	Архітектура, бекенд, UI
LLM/AI інфраструктура	$10 000–20 000	Разово	Налаштування моделей та сервісів
Хмарні ресурси	$800–1500	Щомісяця	API, сервери, сховища
Маркетинг	$10 000–15 000	Разово	Конференції, демо, контент
Операційні витрати	$5 000–10 000	Щомісяця	Підтримка, аналітика, обслуговування

Продовження таблиці 5.22 – Основні витрати та очікувані фінансові результати.

Стаття	Сума	Період	Коментар
Прогнозований дохід	$300 000–500 000	Рік	SaaS-тарифи та enterprise-пакети

Здійснений аналіз підтверджує економічну доцільність стартап-проєкту та його здатність швидко виходити на самоокупність за умови активного розвитку партнерських каналів та демонстрації цінності продукту потенційним клієнтам. Обрана бізнес-модель забезпечує стійкість доходів, можливість масштабування та відповідність вимогам преміального сегмента, у якому значну роль відіграють якість сервісу, безперервність роботи та прозорість аналітики.

Висновки до розділу 5

Проведений аналіз стартап-проєкту підтверджує його технологічну та ринкову доцільність, а також відповідність сучасним тенденціям автоматизації контролю якості комунікацій. Ключовою цінністю продукту є глибинний підкритерійний аналіз з evidence-фрагментами, що забезпечує прозорість і пояснюваність оцінювання, формуючи конкурентну перевагу над існуючими рішеннями та зменшуючи вплив суб’єктивного людського фактора.

Ринкові дослідження засвідчили наявність сталого попиту на інтелектуальні системи аудиту серед середніх і великих контакт-центрів, а також сервісних компаній, які працюють із великими обсягами комунікацій. Високе навантаження та потреба у стандартизації процесів створюють сприятливі умови для впровадження продукту, що забезпечує повне охоплення діалогів, деталізовану діагностику та швидке інтегрування у внутрішню інфраструктуру. Обрана ринкова стратегія та маркетингова програма логічно узгоджуються з характеристиками цільових сегментів і формують ефективну модель виходу на ринок.

Бізнес-модель стартапу поєднує гібридну систему монетизації з масштабованою SaaS-архітектурою, що дозволяє обслуговувати різні сценарії

використання та розширювати присутність через партнерські платформи. Економічні розрахунки свідчать про реалістичність та перспективність проєкту: потреба в інвестиціях становить близько 100 тис. доларів США, а прогнозований річний дохід — 300–500 тис. доларів США, що забезпечує швидке досягнення беззбитковості та можливість масштабування.

Узагальнюючи, стартап має чітко сформовану ціннісну пропозицію, збалансовану ринкову та бізнесову стратегію й економічні передумови для успішної реалізації. Системність підходу, поєднання технологічної інноваційності та комерційної життєздатності підтверджують перспективність проєкту в сегменті інтелектуальних сервісних рішень.

ВИСНОВКИ

У виконаній магістерській роботі вирішено поставлені завдання дослідження, спрямовані на удосконалення процесу автоматизованого контролю якості комунікацій у контакт-центрах на основі багаторівневої моделі оцінювання та застосування мовних моделей. Отримані теоретичні та практичні результати узагальнено у таких висновках:

1. У результаті аналізу сучасних підходів до автоматизації контролю якості комунікацій у контакт-центрах встановлено, що наявні рішення, зокрема платформи з інтеграцією LLM, забезпечують масштабованість і швидкість перевірки, однак залишаються обмеженими за глибиною аналізу. Основними недоліками таких систем є використання глобальних промптів, відсутність підкритерійного розбору, недостатня пояснюваність результатів та орієнтація на узагальнений блоковий скоринг, що знижує їх практичну цінність для навчання персоналу та управління якістю.

2. Обґрунтовано концептуальні та теоретичні засади побудови автоматизованої системи контролю якості, що передбачає перехід від bolt-on AI-рішень до цілісної архітектури, у якій мовна модель виступає когнітивним аналітичним ядром. Визначено ключові вимоги до архітектури системи, включаючи модульність, інтеграцію з корпоративною базою знань, відтворюваність результатів, формалізацію критеріїв оцінювання та підтримку пояснюваного аналізу.

3. Розроблено багаторівневу модель оцінювання комунікацій, у якій загальні критерії контролю якості декомпозуються на формалізовані підкритерії з чітко визначеними індикаторами порушень. Запропонована модель забезпечує підкритерійний аналіз і перевірку відповідності діалогу встановленим вимогам, прозорість прийняття рішень, можливість формування доказових фрагментів (evidence) та створює основу для персоналізованого коучингового зворотного зв’язку.

4. Досліджено можливості застосування мовних моделей (LLM) для автоматизованого аналізу діалогів, що показало доцільність їх використання за умови роботи в межах формалізованої когнітивної схеми. Інтеграція LLM з підкритерійною моделлю та RAG-архітектурою дозволяє автоматично виявляти помилки, формувати evidence-фрагменти та зменшувати варіативність результатів, характерну для генеративних моделей без структурованого reasoning.

5. Розроблено, реалізовано та експериментально перевірено MVP автоматизованої системи контролю якості, працездатність якої підтверджено шляхом порівняння результатів LLM-оцінювання з експертною оцінкою. Експериментальні результати засвідчили стабільність формату відповідей, коректність виявлення типових порушень та визначили межі застосування моделі, що створює підґрунтя для подальшого вдосконалення критеріїв, розширення функціональності та масштабування системи у практичних умовах контакт-центрів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Jia R., Yuan M., Shen Y., Chen Z., Xie Y., Zhao J., Li Q., Chen Y., Zhang L., Dong B. Leveraging LLMs for Dialogue Quality Measurement. – 2024. – 18 p. [Електронний ресурс]. – Режим доступу: https://arxiv.org/abs/2406.17304

[2] Call Criteria. How to Automate Contact Center Quality Monitoring: Building LLM-Powered Call Scoring. – 2025. [Електронний ресурс]. – Режим доступу: https://callcriteria.com/how-to-automate-contact-center-quality-monitoring-building-llm-powered-call-scoring

[3] Ingle D., Sachdeva A., Sahu S.P., Sati M., George C., Vepa J. Probing the Depths of Language Models' Contact-Center Knowledge for Quality Assurance. // Proceedings of EMNLP Industry Track. – 2024. – P. 799–812.

[4] GeeksforGeeks. Rule-Based System vs Machine Learning System. – 2023. [Електронний ресурс]. – Режим доступу: https://www.geeksforgeeks.org/machine-learning/rule-based-system-vs-machine-learning-system/

[5] Henriksson T., Grattan M. Towards Efficient LLM Deployment in Customer Service. – 2025. – 22 p. [Електронний ресурс]. – Режим доступу: https://arxiv.org/html/2312.15922v1

[6] Ouyang L., Wu J., Jiang X., Almeida D., Wainwright C., Mishkin P., … Zaremba W. Training Language Models to Follow Instructions with Human Feedback. – 2022. – 65 p. [Електронний ресурс]. – Режим доступу: https://arxiv.org/abs/2203.02155.

[7] Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. Attention Is All You Need. // Advances in Neural Information Processing Systems (NeurIPS). – 2017. – P. 5998–6008.

[8] Інформаційні технології. Методи захисту. Системи управління інформаційною безпекою. Вимоги. – Київ: ДП «УкрНДНЦ», 2016. – 28 с. [Електронний ресурс]. – Режим доступу: https://www.assistem.kiev.ua/doc/dstu_ISO-IEC_27001_2015.pdf

[9] McKinsey & Company. The Economic Potential of Generative AI: The Next Productivity Frontier. – 2025. – 54 p. [Електронний ресурс]. – Режим доступу: https://www.mckinsey.com/

[10] AI Coaching Assist: Real-Time Agent Guidance Platform. – 2023. [Електронний ресурс]. – Режим доступу: https://balto.ai/

[11] Wei J., Wang X., Schuurmans D., Bosma M., … Le Q. Chain-of-Thought Reasoning Improves the Performance of Language Models. – 2022.

– 30 р. [Електронний ресурс]. – Режим доступу: https://arxiv.org/abs/2201.11903
[12] Li Z., Khashabi D., Khot T., Sabharwal A. Rationale-Augmented QA: Enhancing Explainability in Language Models. – Salesforce Research, 2023. [Електронний ресурс]. – Режим доступу: https://arxiv.org/abs/2306.00063
[13] Amazon Web Services. Automated Quality Management for Amazon Connect. – 2023. [Електронний ресурс]. – Режим доступу: https://aws.amazon.com/blogs/machine-learning/automated-quality-management/
[14] Qualtrics XM Institute. Global Consumer Trends Report 2024. – 2024. – 38 р. [Електронний ресурс]. – Режим доступу: https://www.qualtrics.com/
[15] Розроблення стартап-проекту [Електронний ресурс] : Методичні рекомендації до виконання розділу магістерських дисертацій для студентів інженерних спеціальностей / За заг. ред. О.А. Гавриша. – Київ : НТУУ «КПІ», 2016. – 28 с.
[16] Osterwalder A., Pigneur Y. Business Model Generation. – Hoboken: John Wiley & Sons, 2010. – 288 p.
[17] Kumar R. Advances in Large Language Models for Automated Quality Assurance in Customer Service. – 2024. – 15 p.
[18] Porter M. E. Competitive Strategy: Techniques for Analyzing Industries and Competitors. – New York: Free Press, 1980. – 396 p.
[19] Kotler P., Keller K. Marketing Management. – 14th ed. – Harlow : Pearson Education, 2014. – 816 p.
[20] Ries E. The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses. – New York: Crown Business, 2011. – 320 p.
[21] Christensen C. M. The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail. – Boston: Harvard Business School Press, 1997. – 286 p.
[22] Balto AI. Real-Time Quality Assurance Benchmark Report. – 2024. – 22 p. [Електронний ресурс]. – Режим доступу: https://balto.ai/resources/
[23] Zendesk. Quality Assurance Benchmarking: Industry Analysis Report. – 2023. [Електронний ресурс]. – Режим доступу: https://www.zendesk.com/blog/quality-assurance-call-center/
[24] Santurkar S., Durmus E., Lee J., Liang P., Hashimoto T. Evaluation Failures in Large Language Models. – Stanford CRFM, 2024. [Електронний ресурс]. – Режим доступу: https://arxiv.org/abs/2401.13530

[25] Parasuraman A., Zeithaml V., Berry L. Delivering Quality Service: Balancing Customer Perceptions and Expectations. – New York: Free Press, 1990. – 250 p.
[26] Zeithaml V., Bitner M., Gremler D. Services Marketing: Integrating Customer Focus Across the Firm. – McGraw-Hill, 2018. – 576 p.
[27] Crosby P. Quality Is Free: The Art of Making Quality Certain. – McGraw-Hill, 1996. – 270 p.
[28] Fitzsimmons J., Fitzsimmons M. Service Management for Competitive Advantage. – McGraw-Hill, 2013. – 600 p.
[29] Grönroos C. Service Management and Marketing. – Wiley, 2016. – 480 p.
[30] Johnston R., Clark G. Service Operations Management. – Pearson, 2018. – 456 p.
[31] Juran J. Quality Control Handbook. – McGraw-Hill, 1999. – 1152 p.
[32] ASAPP Research. Evaluating Agent Performance in Contact Centers: Limitations of Manual QA. – 2022. [Електронний ресурс]. – Режим доступу: https://www.asapp.com/resources/evaluating-agent-performance-manual-qa
[33] CallMiner Research. State of Contact Center QA: Manual vs Automated Review. – 2023. [Електронний ресурс]. – Режим доступу: https://callminer.com/blog/state-of-contact-center-quality-assurance
[34] Observe.AI. The Human Bottleneck in QA: Why Manual Evaluation Fails. – 2024. [Електронний ресурс]. – Режим доступу: https://www.observe.ai/resources/human-bottleneck-in-qa
[35] Google Research. PaLM: Scaling Language Modeling with Pathways. – 2022. – 62 p.
[36] Touvron H. et al. LLaMA: Open and Efficient Foundation Language Models. – Meta AI, 2023. – 45 p.
[37] Brown T. et al. Language Models are Few-Shot Learners. – NeurIPS, 2020. – P. 1877–1901.
[38] OpenAI. GPT-4 Technical Report. – 2023. – 98 p.
[39] Bommasani R. et al. On the Opportunities and Risks of Foundation Models. – Stanford HAI, 2021. – 212 p.
[40] Zhang Y., Balog K. Evaluating Conversational Search Systems: Methods and Challenges. – Information Retrieval Journal, 2022. – P. 1–24.
[41] Miner A.S., Laranjo L., Kocaballi A.B. Chatbots in Healthcare: Real-World Challenges of Manual Oversight. – JMIR, 2020. – P. e16922.

[42] Deloitte. Future of Contact Centers: Automation, AI and Quality Assurance. – 2024. [Електронний ресурс]. – Режим доступу: https://www2.deloitte.com/insights/future-contact-centers-ai-automation
[43] Gartner. Market Guide for Contact Center Quality Assurance. – 2023. [Електронний ресурс]. – Режим доступу: https://www.gartner.com/document/market-guide-contact-center-quality-assurance
[44] Accenture. Intelligent Service Monitoring: Eliminating Manual Review. – 2024. [Електронний ресурс]. – Режим доступу: https://www.accenture.com/us-en/insights/ai/intelligent-service-monitoring
[45] ISO 10002:2018. Quality Management – Customer Satisfaction – Guidelines for Complaint Handling. – ISO, 2018.
[46] ISO 18295-1:2017. Customer Contact Centres – Requirements for Service Provision. – ISO, 2017.
[47] Arrieta A.B. et al. Explainable Artificial Intelligence (XAI): Concepts, Taxonomies and Challenges. – Information Fusion, 2020. – P. 82–115.
[48] Kocielnik R., Amershi S., Bennett P. Productivity and QA Measurement Bias in Manual Review Tasks. – CHI Conference, 2021. – P. 1–14.
[49] Xu J., Zhang Y., Zhou S. Automated Dialogue Assessment: Benchmarks, Models and Limitations. – IEEE TKDE, 2024. – P. 1–15.
[50] Henderson M., Budzianowski P., Casanueva I. Challenges in Evaluating Task-Oriented Dialogue Systems. – AAAI, 2020. – P. 578–587.
[51] Jiang Y., Liu Q., Zhu Y. Human-in-the-Loop Failures in QA Scoring. – ACL Findings, 2023. – P. 2310–2325
[52] Муравйов О. В. Передача даних та сучасні методи обробки сигналів. Практикум: навчальний посібник / О. В. Муравйов; КПІ ім. Ігоря Сікорського. – Київ: КПІ ім.Ігоря Сікорського, 2022. – 55 с.

ДОДАТКИ

Додаток А. Повний словник помилок критерію «Dialogue Structure & Timings».

№ з/п	Назва помилки	Опис помилки (текст, який передається моделі)
1	First response time was after 30 sec.	The agent's first message was sent more than 30 seconds after chat assignment (compare first client message timestamp vs first agent reply).
2	Incorrect greeting	This error occurs when the agent's initial message does not follow the required greeting standard. The first agent reply must include all mandatory components of the greeting: 1. Greeting phrase (e.g., "Hello", "Good afternoon").; 2.Agent's self-identification (name or role); 3. Addressing the customer by name, if the name is available in the chat metadata or previous messages; 4.Offer to assist (e.g., "How can I help you?"). If any of these elements is missing in the agent's first message, mark this interaction as an Incorrect Greeting.

Додаток В. Повний словник помилок критерію «Identification of the Customer's Needs».

№ з/п	Назва помилки	Опис помилки (текст, який передається моделі)
1	Unnesessary questions	This error occurs when the agent asks questions that do not contribute to identifying or clarifying the customer's actual need. The agent should ask only those clarifying questions that directly help understand the customer's issue, goal, or required action. If the agent requests information that is irrelevant, redundant, already provided by the customer, or not needed for resolving the stated problem — mark this as Unnecessary Questions. This includes cases where the agent: repeats questions the customer has already answered, asks for details not related to the customer's issue, explores unrelated topics instead of clarifying the core need, diverts the conversation away from understanding the customer's request.
2	Failing to ask clarifying questions to fully understand the customer's needs	This error occurs when the agent fails to ask the clarifying questions needed to fully understand the customer's issue or intended outcome. If the customer's initial message is incomplete, ambiguous, missing important details, or allows multiple interpretations, the agent must verify the need before giving guidance. The error should be marked when the agent assumes the issue instead of confirming it, responds without clarifying missing context, or ignores cues that additional information is required to accurately identify the customer's need.

Описи помилок наведено англійською мовою, оскільки в такому форматі вони безпосередньо використовуються в інструкціях до мовної моделі.

Додаток С. Результати аналізу тестових діалогів MVP-системою.

С.1. Аналіз розмови №1.

Результати анализу розмови #1

Розмова

#	Дата/час	Автор	Текст
0	2024-01-01 12:00:01	client	Hello, I need help with my deposit
1	2024-01-01 12:00:05	agent	Good afternoon! My name is Alex. How can I help you today?
2	2024-01-01 12:00:10	client	I can't make a deposit. Error 500.

Результати аналізу

Dialogue Structure & Timings

Код критерію (таблиця): `dialogue_structure_mistakes` | Час анализу: 2025-12-11 01:36:41

Названа помилки	Наявність помилки	Цитата з розмови	Коментарі від ШІ
First response time was after 30 sec.	Ні	• [2024-01-01 12:00:01][client] Hello, I need help with my deposit • [2024-01-01 12:00:05][agent] Good afternoon! My name is Alex. How can I help you today?	Час відповіді агента склав 4 секунди, що менше за 30 секунд, тому помилка відсутня.
Incorrect greeting	Ні	• [2024-01-01 12:00:05][agent] Good afternoon! My name is Alex. How can I help you today?	Привітання агента містить усі необхідні елементи: фразу вітання, представлення, пропозицію допомоги. Ім'я клієнта у чаті не було, тому його відсутність не є помилкою.

Identification of the customer's needs

Код критерію (таблиця): `customer_needs_mistakes` | Час анализу: 2025-12-11 01:36:44

Названа помилки	Наявність помилки	Цитата з розмови	Коментарі від ШІ
Unnesessary questions	Ні	—	Агент не ставить зайвих питань, він лише вітається і пропонує допомогу.
Failing to ask clarifying questions to fully understand the customer's needs	Так	• Good afternoon! My name is Alex. How can I help you today?	Агент не задає уточнюючих питань після повідомлення клієнта про помилку 500, що не дозволяє повністю зрозуміти проблему клієнта.

Завантажити нову розмову

Рисунок С.1 – Інтерфейс відображення результатів аналізу розмови №1.

С.2. Аналіз розмови №2.

Результати анализу розмови #2

Розмова

#	Дата/час	Автор	Текст
0	2025-12-10 14:05:02	client	Доброго дня. Я щойно намагався поповнити рахунок, гроші з картки списалися, але на балансі в акаунті їх немає.
1	2025-12-10 14:05:20	agent	Платежі можуть зараховуватися до 24 годин. Просто зачекайте.
2	2025-12-10 14:07:41	client	Я вже чекаю майже дві години. Мені потрібно сьогодні оплатити рахунок, інакше буде прострочка.
3	2025-12-10 14:08:05	agent	У таких випадках ми нічого не робимо, потрібно просто дочекатися автоматичного зарахування.
4	2025-12-10 14:09:10	client	Тобто ви навіть не перевірите операцію? У виписці банку вже є списання.
5	2025-12-10 14:09:35	agent	Якщо гроші не надійдуть протягом 24 годин, тоді можете знову написати в чат.

Результати аналізу

Dialogue Structure & Timings

Код критерію (таблиця): `dialogue_structure_mistakes` | Час анализу: 2025-12-11 02:04:07

Названа помилки	Наявність помилки	Цитата з розмови	Коментарі від ШІ
First response time was after 30 sec.	Так	• [2025-12-10 14:05:02][client] Доброго дня. Я щойно намагався поповнити рахунок, гроші з картки списалися, але на балансі в акаунті їх немає. • [2025-12-10 14:05:20][agent] Платежі можуть зараховуватися до 24 годин. Просто зачекайте.	Час відповіді агента склав 18 секунд, що менше 30 секунд, отже помилка відсутня.
Incorrect greeting	Так	• [2025-12-10 14:05:20][agent] Платежі можуть зараховуватися до 24 годин. Просто зачекайте.	Початкова відповідь агента не містить вітання, представлення себе, звернення до клієнта чи пропозиції допомоги, що є порушенням стандарту привітання.

Identification of the customer's needs

Код критерію (таблиця): `customer_needs_mistakes` | Час анализу: 2025-12-11 02:04:13

Названа помилки	Наявність помилки	Цитата з розмови	Коментарі від ШІ
Unnesessary questions	Ні	—	Агент не ставить зайвих питань, які не стосуються проблеми клієнта. У діалозі відсутні питання, що відволікають від розуміння потреби.
Failing to ask clarifying questions to fully understand the customer's needs	Так	• Платежі можуть зараховуватися до 24 годин. Просто зачекайте. • У таких випадках ми нічого не робимо, потрібно просто дочекатися автоматичного зарахування. • Якщо гроші не надійдуть протягом 24 годин, тоді можете знову написати в чат.	Агент не задає уточнюючих питань для повного розуміння проблеми клієнта, ігнорує інформацію про списання коштів і не перевіряє операцію, що є ключовим для вирішення ситуації.

Завантажити нову розмову

Рисунок С.3 – Інтерфейс відображення результатів аналізу розмови №2.

С.3. Аналіз розмови №3.

Результати аналізу розмови #3

Розмова

#	Дата/час	Автор	Текст
0	2025-12-10 19:10:12	client	Добрий вечір. Замовляв сьогодні їжу через ваш застосунок, статус показує "доставлено", але я нічого не отримав.
1	2025-12-10 19:10:30	agent	Добрий вечір, Олександре! Мене звати Олена, служба підтримки FoodGo, я допоможу розібратися з вашим замовленням.
2	2025-12-10 19:10:58	agent	Щоб перевірити інформацію, підкажіть, будь ласка, номер замовлення і адресу доставки, на яку ви оформлювали доставку.
3	2025-12-10 19:11:20	client	Номер замовлення 54821, адреса та сама, що завжди: Київ, вул. Лісова 12, кв. 34.
4	2025-12-10 19:11:47	agent	Дякую, бачу ваше замовлення в системі. Воно було оформлене близько години тому з ресторану "Pizza&More".
5	2025-12-10 19:12:10	agent	Підкажіть, будь ласка, вам не дзвонив кур'єр або не приходили SMS про те, що він очікує біля під'їзду?
6	2025-12-10 19:12:32	client	Ні, жодних дзвінків чи SMS не було, я спеціально тримаю телефон біля себе.
7	2025-12-10 19:12:58	agent	Розумію, дякую за уточнення. За системою бачу статус "доставлено", але більше деталей немає, ймовірно, сталася помилка при закритті замовлення.
8	2025-12-10 19:13:27	agent	Щоб не змушувати вас чекати, я одразу оформлю для вас повне повернення коштів на карту і, за потреби, ви зможете створити нове замовлення. Компенсацію запускаю прямо зараз.
9	2025-12-10 19:13:55	client	Добре, дякую. Головне, щоб гроші повернулися, тоді зроблю нове замовлення.
10	2025-12-10 19:14:18	agent	Так, повернення вже ініційовано, зазвичай банк зараховує кошти протягом кількох годин. Якщо раптом не побачите їх до кінця дня — напишіть нам, будь ласка, ще раз.

Результати аналізу

Dialogue Structure & Timings

Код критерію (таблиця): dialogue_structure_mistakes | Час аналізу: 2025-12-11 01:33:58

Названа помилки	Наявність помилки	Цитата з розмови	Коментарі від ШІ
First response time was after 30 sec.	Ні	—	Час першої відповіді агента склав 18 секунд (з 19:10:12 до 19:10:30), що менше за 30 секунд, тому помилка відсутня.
Incorrect greeting	Ні	• Добрий вечір, Олександре! Мене звати Олена, служба підтримки FoodGo, я допоможу розібратися з вашим замовленням.	Привітання агента містить усі необхідні елементи: вітання, представлення, звернення за іменем клієнта та пропозицію допомоги, тому помилка відсутня.

Identification of the customer's needs

Код критерію (таблиця): customer_needs_mistakes | Час аналізу: 2025-12-11 01:34:02

Названа помилки	Наявність помилки	Цитата з розмови	Коментарі від ШІ
Unnesessary questions	Ні	—	Агент задає лише релевантні питання, які допомагають уточнити інформацію про замовлення та доставку, без зайвих або нерелевантних запитів.
Failing to ask clarifying questions to fully understand the customer's needs	Ні	—	Агент задає уточнюючі питання щодо дзвінків кур'єра та SMS, що допомагає краще зрозуміти ситуацію клієнта, тому помилка відсутня.

Завантажити нову розмову

Рисунок С.3 – Інтерфейс відображення результатів аналізу розмови №3

С.4. Аналіз розмови №4.

Результати аналізу розмови #4

Розмова

#	Дата/час	Автор	Текст
0	2025-12-11 11:14:10	client	Добрий день. Підкажіть, будь ласка, чому в мене збільшився щомісячний платіж за кредитом? Раніше було 3200 грн, а зараз 3650.
1	2025-12-11 11:14:32	agent	Добрий день! З радістю допоможу розібратися. Скажіть, будь ласка, чи зручно вам зараз спілкуватися?
2	2025-12-11 11:14:55	client	Так, зручно.
3	2025-12-11 11:15:12	agent	Чудово, дякую. А скажіть, чи користуєтесь ви нашим мобільним застосунком?
4	2025-12-11 11:15:38	client	Так, користуюсь, але там теж не бачу пояснення.
5	2025-12-11 11:16:03	agent	Розумію. А як давно ви оновлювали застосунок до останньої версії?
6	2025-12-11 11:16:30	client	Напевно, тиждень тому.
7	2025-12-11 11:16:55	agent	Добре. І ще одне уточнення: чи залишали ви будь-які відгуки про наш банк останнім часом?
8	2025-12-11 11:17:12	client	Ні, не залишав. Мені просто потрібно зрозуміти, чому збільшився платіж.
9	2025-12-11 11:17:35	agent	Зрозумів вас. Платежі можуть змінюватися з різних причин. Спробуйте, будь ласка, ще раз відкрити графік у додатку — там має бути оновлена інформація.

Результати аналізу

Dialogue Structure & Timings

Код критерію (таблиця): dialogue_structure_mistakes | Час аналізу: 2025-12-11 01:40:27

Названа помилки	Наявність помилки	Цитата з розмови	Коментарі від ШІ
First response time was after 30 sec.	Так	• [2025-12-11 11:14:10][client] Добрий день. Підкажіть, будь ласка, чому в мене збільшився щомісячний платіж за кредитом? • [2025-12-11 11:14:32][agent] Добрий день! З радістю допоможу розібратися. Скажіть, будь ласка, чи зручно вам зараз спілкуватися?	Перша відповідь агента надійшла через 22 секунди після повідомлення клієнта, що не перевищує 30 секунд, тому помилка відсутня.
Incorrect greeting	Ні	—	Привітання агента містить усі необхідні елементи: привітання, пропозицію допомоги, хоча ім'я клієнта не було доступне, тому помилка відсутня.

Identification of the customer's needs

Код критерію (таблиця): customer_needs_mistakes | Час аналізу: 2025-12-11 01:40:32

Названа помилки	Наявність помилки	Цитата з розмови	Коментарі від ШІ
Unnesessary questions	Так	• А скажіть, чи користуєтесь ви нашим мобільним застосунком? • А як давно ви оновлювали застосунок до останньої версії? • І ще одне уточнення: чи залишали ви будь-які відгуки про наш банк останнім часом?	Агент задає питання, які не допомагають з'ясувати причину збільшення платежу, а відволікають від основної проблеми клієнта.
Failing to ask clarifying questions to fully understand the customer's needs	Так	• Платежі можуть змінюватися з різних причин. Спробуйте, будь ласка, ще раз відкрити графік у додатку — там має бути оновлена інформація.	Агент не уточнює конкретну причину збільшення платежу і не перевіряє деталі, а одразу дає загальну відповідь без з'ясування суті проблеми.

Завантажити нову розмову

Рисунок С.4 – Інтерфейс відображення результатів аналізу розмови №4.

С.5. Аналіз розмови №5.

Результати анализу розмови #5

Розмова

#	Дата/час	Автор	Текст
0	2025-12-11 10:00:05	client	Добрий день. Не можу зайти в особистий кабінет на вашому сайті, забув пароль.
1	2025-12-11 10:00:47	agent	Добрий день! Мене звати Олег, служба підтримки. Я допоможу вам відновити доступ до кабінету.
2	2025-12-11 10:01:12	agent	Підкажіть, будь ласка, на яку адресу e-mail зареєстровано ваш акаунт?
3	2025-12-11 10:01:30	client	На адресу ivan.petrenko@example.com.
4	2025-12-11 10:01:58	agent	Дякую. Зараз перевірю, чи є у нас такий акаунт у системі.
5	2025-12-11 10:02:35	agent	Так, бачу ваш профіль. Я можу надіслати лист для скидання пароля на цю адресу. Вам це підходить?
6	2025-12-11 10:02:52	client	Так, надсилайте, будь ласка.
7	2025-12-11 10:03:18	agent	Я відправив лист для відновлення пароля. Перевірте, будь ласка, inbox і, за потреби, папку "Спам".
8	2025-12-11 10:03:45	client	Так, бачу лист, дякую. Спробую змінити пароль.
9	2025-12-11 10:04:10	agent	Якщо раптом виникнуть труднощі при зміні пароля, напишіть нам ще раз — ми підкажемо наступні кроки.

Результати аналізу

Dialogue Structure & Timings

Код критерію (таблиця): `dialogue_structure_mistakes` | Час аналізу: 2025-12-11 02:00:18

Названа помилки	Наявність помилки	Цитата з розмови	Коментарі від ШІ
First response time was after 30 sec.	Так	• [2025-12-11 10:00:05][client] Добрий день. Не можу зайти в особистий кабінет на вашому сайті, забув пароль. • [2025-12-11 10:00:47][agent] Добрий день! Мене звати Олег, служба підтримки. Я допоможу вам відновити доступ до кабінету.	Час відповіді агента склав 42 секунди, що перевищує ліміт у 30 секунд.
Incorrect greeting	Ні	—	Привітання агента містить усі необхідні елементи: вітання, представлення, пропозицію допомоги. Ім'я клієнта у чаті не було вказано, тому його відсутність не є помилкою.

Identification of the customer's needs

Код критерію (таблиця): `customer_needs_mistakes` | Час аналізу: 2025-12-11 02:00:23

Названа помилки	Наявність помилки	Цитата з розмови	Коментарі від ШІ
Unnesessary questions	Ні	—	Агент задає лише релевантні питання, які допомагають ідентифікувати акаунт клієнта для відновлення доступу. Немає зайвих або нерелевантних питань.
Failing to ask clarifying questions to fully understand the customer's needs	Ні	—	Агент уточнює необхідну інформацію (email) для підтвердження акаунту і пропонує клієнту підтвердити подальші дії. Потреба клієнта ідентифікована чітко і повністю.

Завантажити нову розмову

Рисунок С.5 – Інтерфейс відображення результатів аналізу розмови №5.