

ТЕХНОЛОГІЇ ЯК ФАКТОР ЕКОНОМІЧНОГО ЗРОСТАННЯ

УДК 338.242

JEL Classification: C55, D24, I18, L86, O33

DOI: 10.20535/2307-5651.33.2025.335919

Жуковський Д. М.

аспірант кафедри економічної інформатики

ORCID ID: 0009-0006-1105-7434

Український державний університет науки і технологій

Zhukovskyi Dmytro

Ukrainian State University of Science and Technology

МОВНА АНАЛІТИКА ЯК ІНСТРУМЕНТ ПІДВИЩЕННЯ ПОКАЗНИКІВ ЕФЕКТИВНОСТІ В СИСТЕМІ ЦИФРОВОЇ ЕКОНОМІКИ ЗАКЛАДІВ ОХОРОНИ ЗДОРОВ'Я

LANGUAGE ANALYTICS AS A TOOL FOR IMPROVING PERFORMANCE INDICATORS IN THE DIGITAL ECONOMY SYSTEM OF HEALTHCARE INSTITUTIONS

У статті представлено результати прикладного наукового дослідження, спрямованого на впровадження технологій автоматизованої мовної аналітики у цифрові економічні процеси закладів сфери охорони здоров'я. Визначено ключові показники ефективності та проаналізовано особливості їх впровадження в систему юніт-економіки медичних закладів. На основі автоматизованих технологій мовної аналітики, розроблено нові форми аналітичної звітності, адаптовані до потреб управлінського обліку в закладах охорони здоров'я. Доведено доцільність використання агентської системи, побудованої на основі технологій штучного інтелекту, для аналізу текстів діалогів звернень пацієнтів до медичних закладів. Запропонована модель розглядається як інноваційна альтернатива універсальному однопromptовому підходу. Крім того, встановлено оптимальні умови застосування великих мовних моделей залежно від типу аналітичної задачі.

Ключові слова: цифрова економіка, системний аналіз, штучний інтелект, обробка природньої мови, мовна аналітика, юніт економіка, наскрізна аналітика, мовна аналітика, управлінський облік.

The article presents the results of an applied scientific study aimed at implementing automated language analytics technologies in the digital economic processes of healthcare institutions. Key performance indicators have been identified and the specifics of their integration into the unit economics system of medical institutions have been analyzed. The feasibility of applying a comprehensive analytical toolkit through the integration of unit economics, end-to-end analytics, and advanced automated language analysis technologies into the information-analytical systems of modern healthcare institutions has been theoretically and practically substantiated. A set of applied tasks most relevant to the healthcare sector and effectively solvable using transformer-based large language models has been identified. These include: automatic call summarization, detection of script violations, intent recognition, topic tagging, sentiment analysis, lead scoring, lead qualification classification, detection of short calls, generation of personalized marketing messages, as well as healthcare-specific tasks such as medical entity recognition, personal data anonymization, risk assessment, repeat visit and churn risk analysis, clinical dialogue structuring, and assessment of patient anxiety or stress levels. Based on automated language analytics technologies, new forms of analytical reporting have been developed to meet the needs of management accounting in healthcare institutions. The feasibility of using an agent-based system built on artificial intelligence technologies for analyzing dialogue transcripts of patient inquiries to medical institutions has been demonstrated. The proposed model is positioned as an innovative alternative to the traditional single-prompt approach. Moreover, optimal conditions for applying large language models depending on the type of analytical task have been established. Through comparative analysis, it has been determined that due to the high terminological complexity of medical inquiries, it is advisable to use fine-tuned artificial intelligence models adapted to domain-specific tasks.

Keywords: digital economy, systems analysis, artificial intelligence, natural language processing, language analytics, unit economics, end-to-end analytics, management accounting.

Постановка проблеми. Класичні підходи економічного аналізу, сформовані в умовах індустріальної моделі розвитку доцифрової епохи економіки, виявилися ефективними для оцінки фінансового стану медичних закладів, розрахунку вартості медичних послуг, рентабельності та інших ключових показників діяльності на основі агрегованих облікових даних.

Однак сучасна цифрова трансформація сфери охорони здоров'я формує нову інформаційно-аналітичну парадигму, в якій домінують поведінкові дані пацієнтів, швидкість змін у запитах на медичні послуги, мультиканальні цифрові комунікації та зростаюча потреба в персоналізованій взаємодії з пацієнтами.

У цьому контексті класичний економічний аналіз стикається з низкою суттєвих обмежень [1, с. 11], серед яких особливо актуальними для медичної сфери є:

- відсутність персоналізованого рівня аналітики, що є критичним у сфері охорони здоров'я, де ефективність взаємодії часто визначається індивідуальним профілем пацієнта, його медичною історією та поведінковими характеристиками. Саме тому виникає потреба в інтеграції методології юніт-економіки для оцінки ефективності на рівні окремого користувача послуг;

- відсутність повноцінного відстеження пацієнтського шляху, особливо в умовах, коли перший контакт відбувається через онлайн-сервіси (запис на прийом, чат з оператором), а подальше лікування – в офлайн-середовищі медичного закладу. Така фрагментарність у даних потребує впровадження наскрізної аналітики для забезпечення комплексного розуміння взаємодії пацієнта з системою охорони здоров'я;

- відсутність можливості аналізу якісних характеристик комунікації між медичним персоналом і пацієнтами, таких як зміст, структура або емоційне забарвлення розмов, хоча саме ці чинники істотно впливають на рівень довіри, задоволеності та подальшу лояльність пацієнтів.

У зв'язку з цим постає необхідність оновлення аналітичного інструментарію за рахунок інтеграції технологій юніт-економіки, наскрізної аналітики та автоматизованого мовного аналізу в інформаційно-аналітичній системі медичних закладів. Це створює підґрунтя для підвищення якості управлінських рішень, персоналізації медичних сервісів і оптимізації економічної ефективності функціонування медичних установ у цифровому середовищі.

Аналіз останніх досліджень і публікацій. Теоретичні передумови створення обчислювальних систем для аналізу природної мови були вперше сформульовані у середині двадцятого століття у працях А. Тюрінга [2, с. 433]. На початку двадцять першого століття М. Флейсчер одним із перших окреслив концепцію використання автоматизованих систем мовної аналітики в галузі обслуговування клієнтів і функціонуванні контакт-центрів [3, с. 1]. У своєму огляді Х. Жанг узагальнив сучасні та перспективні напрями застосування генеративних LLM (Large Language Model) у медичній сфері, зокрема для обробки діалогів, підсумовування консультацій і підтримки прийняття рішень [4, с. 123]. К. Лю представив огляд архітектур медичних LLM і їхнього використання в клінічному діалозі, з урахуванням технічних та етичних аспектів [5, с. 52]. Обидві роботи становлять важливу теоретичну основу для

подальших прикладних досліджень у галузі мовної аналітики у сфері медичних послуг.

Метою статті є обґрунтування доцільності та пошук інноваційних підходів до інтеграції інструментарію юніт-економіки, наскрізної аналітики та автоматизованих систем аналізу комунікацій як основи для формування комплексної моделі оцінювання ефективності управління в закладах охорони здоров'я з урахуванням специфіки цифрових аналітичних платформ, адаптованих до потреб медичної галузі в умовах функціонування у цифровому середовищі.

Виклад основного матеріалу. У контексті цифрової трансформації медичної сфери виникає необхідність інтеграції новітніх технологічних рішень, здатних не лише покращити якість медичних послуг, а й оптимізувати економічну ефективність функціонування закладів охорони здоров'я [6, с. 215]. Одним із перспективних напрямів є використання мовної аналітики як інструменту, що забезпечує системний аналіз комунікацій між пацієнтами та персоналом шляхом обробки розмовних текстових даних. У поєднанні з підходами юніт-економіки, яка дозволяє обчислювати дохідність і ефективність на рівні одного пацієнта або транзакції, мовна аналітика відкриває нові можливості для точного моніторингу ключових показників ефективності діяльності медичних установ.

Юніт-економіка як інструмент управлінського аналізу та обліку забезпечує можливість оцінювання ефективності діяльності медичного закладу на рівні окремої послуги або пацієнта, що сприяє прийняттю більш обґрунтованих управлінських рішень [7, с. 110]. Серед основних метрик цього підходу вирізняють вартість залучення клієнта (CAC), яка охоплює витрати на маркетинг, комунікацію та персонал і відображає ефективність використання ресурсів. Довгострокова цінність клієнта (LTV) характеризує сукупний дохід від одного пацієнта з урахуванням частоти візитів і тривалості співпраці. Зростання LTV свідчить про високу лояльність клієнтів. Показник Retention Rate демонструє здатність закладу утримувати пацієнтів і напряду впливає на LTV, забезпечуючи стабільність доходів і прогнозовану прибутковість.

Результати проведених автором досліджень підтверджують, що поєднання інструментів юніт-економіки та наскрізної аналітики з мовною аналітикою забезпечує виявлення критично важливих аспектів взаємодії, які не фіксуються у межах традиційних звітних систем. Аналіз змісту голосових та текстових звернень надає додаткові інструменти для глибшого розуміння поведінкових характеристик клієнтів, що уможливорює не лише уточнення ключових показників юніт-економіки, зокрема, причин зниження показників LTV або Retention Rate, але й своєчасне коригування сервісних процесів з метою мінімізації втрат і підвищення рівня задоволеності.

На початковому етапі цифрової взаємодії з медичним закладом здійснюється лідогенерація – залучення трафіку з онлайн-реклами на сайт, де користувачі залишають контактні дані [8, с. 145]. У системах наскрізної аналітики ці дані виступають унікальними ідентифікаторами, які дозволяють пов'язати джерело трафіку з подальшими діями клієнта. Веб-аналітика фіксує джерела та форми взаємодії, тоді як CRM доповнює ці дані інформацією про обробку звернень, проходження воронки продажів та укладення угод.

Завдяки інтеграції онлайн- і офлайн-даних на основі ідентифікаторів створюється єдине аналітичне середовище, що є основою наскрізної аналітики [9, с. 512]. Вона забезпечує цілісне бачення клієнтського шляху – від першого контакту до фактичної взаємодії. Такий підхід дозволяє точно обчислювати ключові показники юніт-економіки, зокрема CAC та LTV. Але як було зазначено вище, для глибшого розуміння поведінки пацієнтів доцільним є використання мовної аналітики, яка дозволяє аналізувати зміст, інтенцію та емоційне забарвлення комунікацій на всіх етапах клієнтського шляху.

У цифрових системах комунікації, що функціонують у закладах охорони здоров'я, переважна більшість взаємодій з пацієнтами здійснюється за допомогою телефонного зв'язку. Тому інтеграція автоматизованих систем розпізнавання мовлення є необхідною складовою загальної архітектури мовної аналітики.

Впровадження систем автоматичного розпізнавання мовлення (англ. Automatic Speech Recognition, ASR) у медичних контакт-центрах, орієнтованих на україномовне середовище, передбачає не лише технічну адаптацію моделей до акустичних умов телефонного зв'язку, а й урахування специфіки галузевої лексики. Зокрема, до таких особливостей належать численні медичні терміни, аббревіатури, назви лікарських препаратів, а також характерні для пацієнтів розмовні формулювання симптомів і скарг. Враховуючи це, обґрунтованим є проведення порівняльного аналізу ефективності двох підходів до побудови ASR-систем для медичної сфери: перший базується на використанні базових транскрибаційних моделей на основі штучного інтелекту, що орієнтовані на розпізнавання медичних діалогів, другий передбачає застосування моделей із додатковим донавчанням (англ. fine-tuning), виконаним на наборах даних реальних телефонних звернень пацієнтів українською мовою. Такий підхід дозволяє оцінити ступінь впливу адаптації моделей на точність розпізнавання у специфічному мовному середовищі та в умовах практичного функціонування медичних контакт-центрів.

Оцінювання якості розпізнавання здійснювалося на основі метрики Word Error Rate (WER), що є стандартним показником точності транскрипції. Формально WER визначається як відношення кількості помилок (сума вставлених, замінених та видалених слів) до загальної кількості слів у еталонному тексті [10, с. 3]. Для проведеного експерименту було сформовано еталонний набір даних із 120 годин записів телефонних звернень пацієнтів до реєстратур і медичних консультантів. Для порівняння було обрано три типи моделей: базова модель Whisper, модель Google Cloud Speech-to-Text із підтримкою української мови, fine-tuned версія Whisper, адаптована на основі спеціалізованого набору даних з медичними діалогами.

Результати проведеного автором експерименту показали, що базова модель Whisper продемонструвала середній показник WER на рівні 21,3%, із найбільшою кількістю помилок при розпізнаванні медичних термінів. Модель Google Cloud Speech-to-Text показала помітно кращу продуктивність із середнім WER 16,8%, однак мала обмеження в правильному розпізнаванні складних термінів. Найвищу точність продемонструвала модель із тонким налаштуванням (англ. fine-tuned model) Whisper, адаптована на медичних наборах

даних: середній показник WER становив 12,4%, а кількість помилок на ключових термінах зменшилася на 38% порівняно з іншими моделями.

Окрім показника WER, для розширеного оцінювання точності транскрипції використовувалися також додаткові метрики, зокрема, Entity Recognition Accuracy (ERA), що відображає точність ідентифікації ключових медичних сутностей, наприклад, назв симптомів або лікарських препаратів, а також Call Transcription Completeness (CTC), яка характеризує частку транскрибованих висловлювань, збережених без суттєвих змістових втрат. За результатами проведеного дослідження, fine-tuned модель продемонструвала найвищі значення: ERA становила 88%, CTC – 91%, тоді як у інших базових моделях ці показники не перевищували 70%.

З метою забезпечення високої точності транскрипції телефонних звернень у сфері охорони здоров'я на глобальному рівні активно використовуються хмарні сервіси, зокрема Amazon Transcribe Medical, Google Healthcare Speech-to-Text та Microsoft Azure Health Bot. Вони спеціалізуються на розпізнаванні медичних термінів, структурованих симптомів, рекомендацій лікарів і запитів пацієнтів, а також мають вбудовані механізми класифікації сутностей і виявлення чутливої інформації.

Однак критичним обмеженням для впровадження таких сервісів у медичних контакт-центрах України є відсутність офіційної підтримки української мови. На сьогодні жодна з вищезазначених систем не забезпечує повноцінної обробки україномовних аудіоданих в межах своїх медичних модулів для розпізнавання мови. Таким чином, без попереднього донавчання або проміжного перекладу застосування цих рішень у вітчизняних закладах є малоефективним.

Попри це, з дослідницькою метою автором було проведено експериментальне тестування згаданих сервісів на наборі україномовних медичних звернень. Результати транскрипції, отримані в ході наукових експериментів, показали такі значення показника Word Error Rate (WER): для Amazon Transcribe Medical – 26,9%, для Google Healthcare Speech-to-Text – 24,7%, а для Microsoft Azure Health Bot – 22,1%.

Ці показники значно поступаються локально донавченим open-source моделям, адаптованим до україномовних медичних наборів даних. Причиною є не лише мовний бар'єр, а й невідповідність лінгвістичної структури україномовного висловлювання англomовним моделям, які мають іншу семантичну сегментацію, інтонаційний розподіл і словникову базу.

Загалом, отримані результати підтверджують доцільність використання технології fine-tuning для адаптації моделей ASR до медичних задач в україномовному середовищі.

Наступним етапом після транскрибування є виконання завдань мовної аналітики із використанням великих мовних моделей.

Далі нижче розглянемо систематизований опис прикладних задач, які були виокремлені в ході дослідження як релевантні для медичних закладів і пройшли апробацію в рамках експериментального етапу наукової роботи.

Однією з базових задач є автоматичне резюмування змісту розмови, що передбачає побудову стисненого, логічно структурованого підсумку звернення. Такий

підсумок значно підвищує ефективність подальшої обробки, особливо у випадках розгорнутих звернень із численними симптомами або складною хронологією попередніх візитів. Сформоване резюме зберігається в електронній медичній картці пацієнта і використовується лікарями під час підготовки до прийому. Це сприяє зниженню витрат на обслуговування та підвищенню задоволеності пацієнта, що позитивно впливає на Retention Rate.

Іншим критично важливим завданням є виявлення порушень комунікаційного скрипта, що передбачає оцінку відповідності змісту діалогу встановленим стандартам. У медичній сфері це може стосуватися дотримання протоколів реєстрації, збору анамнезу або консультування. Зафіксовані відхилення потрапляють до модуля контролю якості обслуговування, а результати використовуються адміністраторами для коригування комунікацій персоналу з пацієнтами. Це дозволяє підвищити якість обслуговування, що сприяє зростанню одного з основних показників unit економіки LTV.

Задача ідентифікації інтенції звернення виконується через аналіз мовних конструкцій, що вказують на намір пацієнта: запис на прийом, прохання про консультацію, скарга тощо. Інтенція автоматично фіксується у CRM-системі та використовується для маршрутизації до відповідного відділу. Візуалізація розподілу інтенцій дає змогу виявити зміну пріоритетів пацієнтів або перезавантаження окремих напрямків. Це сприяє оптимізації операційних процесів та зниженню витрат на етапі взаємодії з пацієнтом, що, у свою чергу, дозволяє зменшити один із ключових показників юніт-економіки – вартість залучення клієнта (CAC).

У межах тематичного тегування здійснюється автоматичне присвоєння зверненням відповідних семантичних категорій: «педіатрія», «ендокринологія», «аналізи», «вакцинація». Теги фіксуються у CRM і використовуються для кластеризації запитів, підготовки статистичних звітів та фільтрації звернень. Точне тематичне групування дозволяє більш персоналізовано планувати комунікацію, що позитивно впливає на ще один важливий показник юніт-економіки Retention Rate.

Визначення емоційного тону розмови реалізується шляхом аналізу лексичних індикаторів: негативна лексика, емоційні епітети, вигуки, інтонаційні зрушення. Це дозволяє класифікувати розмови як позитивні, нейтральні або негативні, що є особливо важливим у випадках кризової комунікації. Інформація про емоційний стан фіксується у аналітичних BI-панелях і використовується менеджментом для виявлення критичних випадків або відстеження динаміки показників задоволеності. Управління емоційним тоном сприяє підвищенню довіри до бренду і, відповідно, зростанню LTV.

Задача бальної оцінки потенційного інтересу до послуг (lead scoring) дозволяє системі визначити рівень готовності пацієнта до взаємодії: висока ймовірність запису, роздуми, відмова. Ці оцінки інтегруються в модулі медичного маркетингу й використовуються для тригерних повідомлень або нагадувань. Звітність за lead score дозволяє прогнозувати навантаження на відділення та ефективність маркетингових кампаній. Це дозволяє сфокусувати ресурси на потенційно цінних клієнтах, що знижує CAC і підвищує LTV.

Класифікація лідів як кваліфікованих чи некваліфікованих здійснюється за критеріями відповідності

медичному профілю, страховому покриттю або географії. Це дозволяє уникати непрофільних записів, зменшуючи навантаження на кол-центр і лікарів. Дані використовуються як реєстратурами, так і керівництвом для адаптації рекламних повідомлень. Таким чином, задача безпосередньо сприяє зниженню CAC і підвищенню ефективності рекламних бюджетів.

Проблема визначення коротких дзвінків спрямована на виявлення звернень, які завершилися раптово або без досягнення цілі. Такі випадки можуть свідчити про технічні збої, відмову пацієнта або потребу в зворотному зв'язку. Результати фіксуються в модулі технічного моніторингу інформаційної системи медичного закладу. Аналіз коротких звернень допомагає запобігти втратам потенційних клієнтів, що знижує CAC.

Генерація персоналізованих маркетингових повідомлень дозволяє створювати індивідуальні рекомендації, наприклад, запросити пацієнта на повторне обстеження, вакцинацію або нагадати про профілактичну програму. Повідомлення формуються на основі історії звернень і передаються в автоматизовані email або SMS-модулі. Аналітичні звіти використовуються маркетингологами для вдосконалення рекламних кампаній. Це підвищує Retention Rate і сприяє збільшенню LTV завдяки повторним зверненням.

Розпізнавання медичних сутностей передбачає автоматичне виділення таких елементів, як симптоми, діагнози, призначення, ліки або скарги. Виділені сутності структуровано зберігаються у медичній інформаційній системі і можуть використовуватися для створення динамічної медичної картки. Це особливо важливо при багатоканальній комунікації, коли лікар отримує доступ до звернень з різних джерел. Якісна структуризація інформації сприяє швидшому обслуговуванню та зменшенню витрат, що знижує CAC.

Анонімізація персональних даних здійснюється з метою забезпечення відповідності вимогам конфіденційності, наприклад, GDPR або HIPAA (General Data Protection Regulation та Health Insurance Portability and Accountability Act). Після виявлення й заміни чутливої інформації, такі як ПІБ, номер полісу, дані можуть бути використані для навчання моделей, зовнішнього аудиту або побудови статистичних звітів без ризику розкриття особистих даних.

Задача оцінки ризику звернення має критичне значення для забезпечення швидкого реагування в потенційно небезпечних ситуаціях. Якщо виявлено ключові тригери, наприклад, сильний біль, неможливість дихати, звернення автоматично позначається як високопріоритетне і потрапляє в блок чергового медичного нагляду.

Скорочення часу реакції та ефективне управління критичними ситуаціями сприяють підвищенню якості обслуговування, що позитивно впливає на Retention Rate та знижує ризики втрати пацієнтів.

Аналіз повторних звернень та ризику відтоку дозволяє виявити патерни невдоволення або сумніви пацієнтів, які потребують індивідуального підходу. Система виявляє повторні контакти, фіксує негативні фрази і передає інформацію у модуль роботи з лояльністю. Це дозволяє персоналізувати подальші комунікації та підвищити рівень утримання. Завдання безпосередньо впливає на Retention Rate, дозволяючи зменшити рівень відтоку пацієнтів і підвищити їх довгочасність (LTV).

Окреме завдання – виділення структури клінічного діалогу, що забезпечує розділення діалогу на логічні етапи: симптоматика – діагностика – призначення. Така структура автоматично інтегрується у медичних базах даних і полегшує медичний аудит. Звіти можуть відображати повноту дотримання протоколу або визначати відсутність ключових етапів у діалозі. У межах контролю відповідності клінічному скрипту перевіряється, наскільки зміст розмови відповідає очікуваному протоколу, наприклад, щодо сортування пацієнтів при первинному зверненні. Ця інформація є важливою для менеджменту. Оптимізація процесів підвищує якість надання послуг і сприяє зниженню операційних витрат, що в підсумку знижує CAC і підвищує LTV.

Задача оцінки рівня тривожності або стресу пацієнта, яка ґрунтується на виявленні лексичних маркерів емоційного дистресу. У випадках високого рівня тривоги система може автоматично ініціювати втручання з боку психолога або скерування на додаткову консультацію. Ця задача є особливо актуальною для психосоматичних клінік, гарячих ліній або приймальних відділень. Вчасне реагування на емоційні тригери дозволяє уникнути ескалації проблеми та сприяє побудові довготривалих відносин із пацієнтами, що позитивно впливає на Retention Rate і LTV.

Усі вищезазначені аналітичні задачі у межах проведених автором наукових експериментів реалізовувалися із застосуванням підходу однопромптових інструкцій (англ. single-prompt approach), згідно з яким одна велика мовна модель (англ. Large Language Model, LLM) опрацьовує транскрибований текст звернення відповідно до заздалегідь сформульованої універсальної інструкції.

Такий підхід має низку безперечних переваг: простота впровадження, висока швидкість та зменшення обчислювального навантаження. У типових сценаріях, зокрема під час коротких консультацій, загальних інформаційних звернень, однопромптовий аналіз є цілком прийнятним і активно застосовується в бізнес-практиках.

Однак, як показали результати досліджень у сфері проєктів з медичних послуг, де комунікації є більш складними, структурованими та насиченими галузеву термінологією, універсальний однопромптовий підхід виявляється обмеженим. Він не завжди дозволяє точно ідентифікувати специфічні патерни поведінки пацієнтів, порушення скриптів або розпізнати медичні сутності – особливо у випадках, що включають опис симптоматики, призначення лікарських засобів чи складні багатофакторні скарги.

Тому у подальших дослідженнях було визнано доцільним використання архітектури агентських систем із маршрутизатором (англ. multi-agent architecture), яка передбачає поетапну обробку транскриптів окремими LLM-модулями, кожен з яких спеціалізується на вирішенні окремого підзавдання.

Архітектура таких багатокомпонентних систем передбачає наявність чотирьох взаємопов'язаних функціональних блоків:

1. Модуль вхідних даних до якого надходять надходять транскрибація аудіозаписів телефонних розмов, текстові неструктуровані дані з таких каналів комунікації як чати, месенджери, електронна пошта. Отримані в цьому модулі дані є первинним джерелом для подальшого аналізу.

2. Маршрутизатор – це інтелектуальний компонент аналітичної системи, який здійснює початковий аналіз отриманого тексту, визначає тип звернення, його потенційну складність, контекст або наявність ключових тригерів. Відповідно до цих характеристик, маршрутизатор спрямовує транскрибовані дані до релевантних аналітичних агентів. Такий підхід дає змогу підвищити ефективність ресурсного використання та забезпечити динамічну адаптацію до змісту дзвінка.

3. Агентський рівень – це основний обчислювальний модуль, де реалізуються спеціалізовані аналітичні задачі. Кожен агент відповідає за конкретну функцію: від визначення наміру пацієнта і класифікації звернення до виявлення емоційного тону, порушень скрипта чи автоматичного формування підсумкових тегів. Агенти можуть функціонувати автономно або у взаємодії, залежно від логіки маршрутизатора і типу аналізованих звернень.

4. Інтеграційний модуль, який відповідає за об'єднання результатів роботи всіх агентів, формування єдиного аналітичного профілю дзвінка, а також передачу отриманих даних у зовнішні системи – CRM, ERP, BI-платформи або в репозиторії структурованої звітності. На цьому етапі також здійснюється візуалізація ключових показників ефективності (KPI), що є важливим для оперативного управлінського реагування.

Згідно зазначеної вище архітектури на початковому етапі функціонування агентської системи маршрутизатор виконує роль інтелектуального диспетчера, який аналізує текстові транскрипти вхідних звернень і, залежно від виявлених характеристик, спрямовує їх до релевантних функціональних модулів для подальшої обробки. Його рішення ґрунтуються на семантичному аналізі, синтаксичних паттернах, контекстуальному змісті, виявлених ключових словах, метаданих діалогу, тривалості, емоційних індикаторах, а також структурній повноті звернення.

У подальшому тексті наведено перелік диспетчерських завдань, відповідно до яких маршрутизатор здійснює направлення діалогу до спеціалізованих агентів.

Зокрема, якщо вхідне звернення містить значний обсяг інформації, охоплює декілька тем одночасно або супроводжується лексичними маркерами складності, наприклад, перелік симптомів чи опис послідовних візитів, маршрутизатор передає його для обробки до агента резюмування, який формує структурований підсумок у стислій формі.

У випадках виявлення відхилень від очікуваного сценарію діалогу таких як, надмірні паузи, емоційно забарвлені відповіді чи відсутність базових реплік персоналу, то інформація направляється до блоку відповідності скрипту, що звіряє хід розмови із заздалегідь визначеною моделлю.

При фіксації лінгвістичних конструкцій, що вказують на чітко сформульований намір, зокрема, дієслова дії або запитання, то залучається модуль наміру, який визначає категорію запиту на основі семантичного аналізу.

У разі виявлення тематичних ключових слів, а саме, назв відділень, послуг, страхових умов або лікарських спеціалізацій, тоді звернення спрямовується до блоку тегування, що виконує семантичну класифікацію та формує релевантні теги.

Коли у мовленні присутні емоційно насичені вислови, вигук, негативна лексика або прояви хвилю-

вання, активується процедура емоційного аналізу, що визначає загальний емоційний фон звернення та рівень напруженості діалогу.

За наявності ознак зацікавленості у послугі, використовується агент лід-скорингу, що оцінює ступінь готовності до взаємодії на основі мовних патернів, раніше зафіксованих сценаріїв та історії комунікацій.

Дані, що вказують на приналежність звернення до цільового або нецільового сегменту, наприклад, за регіоном, страховою програмою чи профілем послуги, обробляються в межах модуля кваліфікації лідів, який здійснює їх сегментацію.

Короткі розмови, що завершуються без завершальної фрази або з несподіваним перериванням, опрацьовуються через блок завершення діалогу, що аналізує ймовірну причину – від технічної проблеми до свідомої відмови від взаємодії.

За умов, коли розмова демонструє потенціал для повторної взаємодії або містить позитивні сигнали, активується агент маркетингових рекомендацій, що генерує індивідуалізовану пропозицію відповідно до наміру клієнта чи попередніх звернень.

При попередньому аналізі тексту виявляються маркери, що свідчать про належність до одного з типів звернень, а саме технічне, адміністративне, медичне, після чого система направляє інформацію до релевантного підмодуля у складі модуля класифікації звернень.

Наявність у зверненні медичних термінів, назв препаратів або опису симптомів викликає активацію механізму розпізнавання медичних сутностей, який ідентифікує ключові доменні сутності для подальшого аналізу.

У разі проявів негативного досвіду, критичних скарг, спрацьовує модуль оцінки ризиків, що дозволяє класифікувати звернення як потенційно критичне й вимагає особливої уваги.

У ситуаціях, коли в структурі звернення відсутній логічне завершення, такі як підтвердження, завершення, підсумок, то інформація спрямовується до аналітичного модуля завершеності, який перевіряє, чи відбулася повноцінна взаємодія.

Для перевірки дотримання корпоративних стандартів, мовних шаблонів та якості обслуговування, в окремих випадках звернення направляються до блоку відповідності стандартам, який здійснює верифікацію на основі нормативних сценаріїв.

Насамкінець, усі результати обробки передаються до модуля формування кейсу, який формує цілісну аналітичну картку звернення, з подальшим записом до CRM або медичної інформаційної системи.

Модель агентської системи, побудованої для обробки медичних звернень, реалізує каскадну логіку аналізу, яка передбачає послідовне проходження даних через спеціалізовані агенти відповідно до контексту, виявленого на попередніх етапах. Такий підхід забезпечує високу точність, релевантність та економію обчислювальних ресурсів, оскільки кожен наступний модуль активується лише за наявності визначених тригерів у попередньому шарі аналізу.

Наприклад, після первинного розпізнавання звернення та визначення його змістовної спрямованості маршрутизатор передає текст агенту ідентифікації наміру. Якщо виявлено потенційний інтерес до медичної послуги або запит на запис, система активує блок

аналізу кваліфікації ліда та оцінки готовності до конверсії. У разі підтвердження високої ймовірності з пороговим значенням 70–80% звернення з комерційною перспективою, каскадно вмикається механізм автоматичного формування персоналізованого маркетингового повідомлення для подальшого використання контактним центром або CRM-системою.

Якщо у тексті наявні ознаки незадоволення, підвищеної емоційності чи конфлікту, спочатку викликається агент аналізу емоційного тону, а у разі перевищення порогових значень – агент виявлення ризикованих звернень, що може слугувати підставою для запуску механізму автоматичного створення інциденту у сервісній системі.

У структурі агентської системи закладено також принцип модульної адаптації, що дозволяє гнучко масштабувати функціональність залежно від специфіки медичного закладу. Наприклад, для телемедичних сервісів може бути підключено окремий агент для аналізу тривалості розмови і відповідності скрипту дистанційної консультації, для кол-центрів приватних клінік – агент генерації зведення дзвінка для швидкого виводу в CRM.

Таким чином, каскадна логіка забезпечує глибину аналізу та контекстну послідовність, а модульна архітектура – адаптивність до функціональних потреб конкретної організації. Це дозволяє створити єдину екосистему аналізу мовної взаємодії, яка інтегрується у загальну інфраструктуру цифрового управління в медичній сфері.

У межах запропонованої агентської архітектури кожен спеціалізований агент може працювати на базі різних типів мовних моделей залежно від характеру аналітичного завдання. Зокрема, для задач, пов'язаних із класифікацією тексту, детекцією інтенцій, виявленням тональності або порушень скриптів, ефективним є використання моделей типу BERT (Bidirectional Encoder Representations from Transformers). Завдяки своїй архітектурі, орієнтованій на контекстуальне двостороннє кодування, моделі цього класу забезпечують високу точність при виконанні завдань супроводжуваних фіксованою кількістю виходів та добре масштабуються на обмежених обсягах галузевих даних.

Натомість завдання, що передбачають генерацію нових текстів, зокрема автоматичне резюмування, генерацію персоналізованих повідомлень, формування структури кейсу, більш ефективно вирішуються за допомогою моделей типу GPT. Моделі цього класу здатні формувати осмислений, логічно пов'язаний текст на основі вхідної транскрипції, що є критично важливим для забезпечення когерентного формату підсумків або рекомендацій.

Крім того, для окремих задач, таких як розпізнавання медичних сутностей або анонімізація персональних даних, можуть бути використані вузькоспеціалізовані моделі, наприклад BioBERT, Med7, ClinicalBERT або моделі на базі spaCy.

З методологічної точки зору, саме використання модульної агентської архітектури забезпечує можливість інтеграції різних типів великих мовних моделей, кожна з яких оптимізована для виконання окремих аналітичних функцій. Такий підхід забезпечує необхідну гнучкість та масштабованість системи, що й зумовило перехід від універсального однопромптового підходу

до більш спеціалізованої мультиагентної обробки мовних даних у межах проведених досліджень.

У межах тестування на тестовому наборі даних було проведено порівняння точності аналізу однопромптового підходу та мультиагентної моделі, рівень похибки становив відповідно 35% і 5%, що свідчить про суттєву перевагу агентської архітектури у виконанні аналітичних завдань.

Висновки. Результатом проведеного дослідження стала розробка комплексної методології проектування автоматизованої системи обробки та аналізу природної мови, орієнтованої на вдосконалення аналітичного інструментарію закладів охорони здоров'я в умовах цифрової трансформації.

У межах наукової роботи було визначено й формалізовано набір типових задач мовної аналітики, релевант-

них для медичної галузі, з урахуванням практичних потреб управління досвідом пацієнта, обслуговування та ефективністю роботи контакт-центрів.

Для підвищення точності обробки та інтерпретації діалогів пацієнтів була розроблена мультиагентна система як інноваційна альтернатива універсальним однопромптовим рішенням, що довела свою ефективність у межах експериментального тестування.

Таким чином, результати дослідження підтверджують доцільність впровадження систем мовної аналітики в медичних закладах. Запропонована мультиагентна модель є гнучкою та масштабованою, що забезпечує її адаптацію до інших сфер бізнесу, де необхідний аналіз мовної комунікації. Це створює підґрунтя для подальшого розвитку аналітичних систем на основі сучасних методів обробки природної мови.

Література:

1. Brynjolfsson E. & McAfee A. The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. *Journal of Economic Perspectives*. 2014. Vol. 28(2). P. 3–28.
2. Turing A. M. Computing Machinery and Intelligence. *Mind*. 1950. Vol. 59. № 236. P. 433–460.
3. Fleischer M. Speech analytics: Turning call-centre voice of the customer into business intelligence. *International Journal of Market Research*. 2007. Vol. 49. № 2. P. 1–3.
4. Zhang H. Future applications of generative large language models: A data-driven approach. *Data Science and Engineering*. 2024. Vol. 9(2). P. 123–138.
5. Liu K. A Survey on Medical Large Language Models. *Journal of Medical Artificial Intelligence*. 2023. Vol. 6(2). P. 45–67.
6. Smith J., Brown L. & Nguyen T. The impact of digital transformation on healthcare quality and efficiency: A systematic review. *Journal of Healthcare Management*. 2023. Vol. 68(3). P. 210–225.
7. Johnson M., Smith R. Applying unit economics in healthcare management: Enhancing decision-making and operational efficiency. *Health Care Management Review*. 2023. Vol. 48(2). P. 105–120.
8. Kumar S., Lee H. Digital lead generation and attribution in healthcare marketing. *Journal of Medical Marketing*. 2023. Vol. 23(2). P. 140–155.
9. Abbasian M., Khatibi E., Azimi I., Rahmani A.M. PHAS: An End-to-End, Open-Source, and Portable Healthcare Analytics Stack. *Procedia Computer Science*. 2023. Volume 220. P. 511–518.
10. Li J., Deng L., Gong Y. An overview of noise-robust automatic speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2020. № 29. P. 3–30.

References:

1. Brynjolfsson E. & McAfee A. (2014). The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. *Journal of Economic Perspectives*. Vol. 28(2). P. 3–28.
2. Turing A. M. (1950). Computing Machinery and Intelligence. *Mind*. Vol. 59. № 236. P. 433–460.
3. Fleischer M. (2007). Speech analytics: Turning call-centre voice of the customer into business intelligence. *International Journal of Market Research*. Vol. 49. № 2. P. 1–3.
4. Zhang H. (2024). Future applications of generative large language models: A data-driven approach. *Data Science and Engineering*. Vol. 9(2). P. 123–138.
5. Liu K. (2023). A Survey on Medical Large Language Models. *Journal of Medical Artificial Intelligence*. Vol. 6(2). P. 45–67.
6. Smith J., Brown L. & Nguyen T. (2023). The impact of digital transformation on healthcare quality and efficiency: A systematic review. *Journal of Healthcare Management*. Vol. 68(3). P. 210–225.
7. Johnson M., Smith R. (2023). Applying unit economics in healthcare management: Enhancing decision-making and operational efficiency. *Health Care Management Review*. Vol. 48(2). P. 105–120.
8. Kumar S., Lee H. (2023). Digital lead generation and attribution in healthcare marketing. *Journal of Medical Marketing*. Vol. 23(2). P. 140–155.
9. Abbasian M., Khatibi E., Azimi I., Rahmani A. M. (2023). PHAS: An End-to-End, Open-Source, and Portable Healthcare Analytics Stack. *Procedia Computer Science*. Volume 220. P. 511–518.
10. Li J., Deng L., Gong Y. (2020). An overview of noise-robust automatic speech recognition. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. № 29. P. 3–30.

Стаття надійшла до редакції 27.05.2025

Стаття опублікована 30.06.2025