

# Модели и алгоритмы обработки медицинских текстов с помощью больших языковых моделей

Гаянова М.М.  
Доцент, к.т.н.

Институт информатики, математики и  
робототехники  
Уфимский университет науки и технологий  
Уфа, Россия  
e-mail: gayanova.mm@ugatu.su

Вульфин А.М.  
Профессор, д.т.н.

Институт информатики, математики и  
робототехники  
Уфимский университет науки и технологий  
Уфа, Россия  
e-mail: vulfin.am@ugatu.su

Любченко К.А.

Уфимский университет науки и технологий  
Уфа, Россия  
e-mail: lyubchenko.kseniya@gmail.com

## Аннотация<sup>1</sup>

Работа посвящена разработке моделей и алгоритмов для обработки медицинских текстов с использованием больших языковых моделей. Разработано программное обеспечение, реализующее предложенные алгоритмы и модели в составе вопрос-ответной системы поддержки принятия решений. Полученные результаты могут быть использованы в различных областях медицинской практики, включая диагностику заболеваний, проведение исследований и разработку новых методов лечения.

**Ключевые слова:** большие языковые модели, расширенная дополненная генерация, поддержка принятия решений, вопрос-ответные системы.

## 1. Введение

Большие объемы накопленных в существующих базах и хранилищах медицинских текстовых данных, включая научные статьи, клинические записи, результаты исследований и диагностические отчеты, могут быть использованы для интеллектуального поиска и извлечения знаний в составе интеллектуальных систем поддержки принятия решений для медицинских работников и исследователей.

Проблемой является необходимость структурирования и эффективного управления

накопленными массивами текстовыми данными для повышения точности и скорости интеллектуального поиска, включая построение вопрос-ответных систем. Перспективным направлением решения данной проблемы является разработка моделей и алгоритмов для обработки медицинских текстов с помощью больших языковых моделей (БЯМ).

Несмотря на существование различных технологий и методов, построение эффективных вопрос-ответных модулей в составе систем поддержки принятия решений для конкретных предметных областей остается по-прежнему сложной задачей в виду следующих факторов:

- разнородность и объем текстовых данных – тексты могут быть различных форматов, что затрудняет их единообразную обработку и анализ;
- сложность и разнообразие вопросов – вопросы могут быть как простыми, требующими извлечения фактов, так и сложными, требующими глубокого анализа и понимания текста;
- объективная оценка правильности ответов – для определения корректности найденных ответов необходимы четкие критерии и методы оценки.

Таким образом, целью работы является повышение эффективности генерации ответа на пользовательский запрос по массиву текстовых документов с возможностью оценки его релевантности.

---

Труды X Международной научной конференции "Информационные технологии интеллектуальной поддержки принятия решений", 12-14 ноября, Уфа-Баку-Чандигарх, 2024

## 2. Анализ проблемы обработки медицинских текстов с помощью больших языковых моделей

Текстовая база данных представляет собой структурированное хранилище, содержащее текстовые документы, метаданные, вопросы и ответы. Основными компонентами таких баз данных являются:

- тексты: основной источник информации, из которого извлекаются ответы;
- вопросы: запросы, на которые требуется найти ответы в текстах;
- ответы: эталонные ответы, используемые для сравнения с найденными системой ответами.

Существует несколько основных категорий вопросов, которые можно задать по тексту:

- фактические вопросы: требующие конкретного ответа;
- описание и объяснение: вопросы, требующие развернутого ответа;
- сравнительные вопросы: вопросы, требующие сравнения;

Для поиска ответов на вопросы в текстах используются различные методы и подходы:

- методы извлечения информации: направлены на выделение релевантной информации из текстов. Примеры таких методов включают регулярные выражения, правила на основе шаблонов и частотные анализаторы;
- машинное обучение: использование обученных моделей для предсказания ответов на вопросы. К таким методам относятся нейронные сети, деревья решений и методы, основанные на статистических моделях;
- поиск на основе семантики: применение семантического анализа для понимания смысла текста и вопросов. Включает использование методов обработки естественного языка, таких как Word2Vec, BERT и другие языковые модели.

Для оценки правильности найденных ответов используются следующие критерии:

- точность: насколько точно найденный ответ соответствует эталонному;
- полнота: охватывает ли найденный ответ все аспекты вопроса;
- релевантность: насколько найденный ответ соответствует контексту вопроса.

Методы оценки включают автоматические метрики, такие как BLEU, ROUGE, и человеческую оценку, когда эксперты оценивают качество ответов по заранее определенным параметрам.

Анализ существующих систем показывает, что современные методы и подходы способны решать задачи поиска ответов с высокой точностью и релевантностью, однако каждая из них требует адаптации и настройки под конкретные задачи и условия использования.

Технология больших языковых моделей (Large Language Models, LLM) стала одним из самых перспективных направлений для исследований и разработок в сфере искусственного интеллекта (ИИ) для обработки естественного языка. Применение LLM позволяет существенно повысить точность решения самых разнообразных задач обработки естественного языка, включая интерпретацию и классификацию, обобщение текстовой информации, создание чат-ботов и иных диалоговых систем, генерацию текстов по запросам и т.д.

В работе [1] представлен обзор перспектив применения больших языковых моделей в здравоохранении. Показано, что применение предварительно обученных LLM позволяет создать прикладные модели для решения узких задач с минимальным набором для дообучения ('few-shot') или даже вовсе без обучающего набора ('zero-shot'). В России проводятся исследования и разработки в области LLM. В частности, модель ruGPT-3.5, насчитывающую 13 млрд параметров, создали в Сбере. Аналогичную большую языковую модель YandexGPT-2 (100 млрд параметров) развивает Яндекс.

Авторы исследования показывают, что применение LLM при создании специализированных ИИ-продуктов для клинического применения и поддержки принятия врачебных решений выглядит более обоснованным подходом, чем создание универсального ИИ. Более того, применение LLM демонстрирует гораздо лучшую эффективность в задачах, где не требуется специальных знаний, или заданиях, которые предоставляются в виде развернутых пользовательских запросов (prompt). Это открывает довольно большие возможности для ускорения внедрения ИИ в медицинскую практику, бросая, по сути, вызов традиционным подходам, применяемым при создании систем поддержки принятия врачебных решений (СППВР).

В работе [2] представлена большая языковая модель PaLM 2, которая обладает улучшенными возможностями многоязычия и логического мышления и более эффективна в вычислениях, чем ее предшественница PaLM.

В работе [3] описана коллекция БЯМ Llama2, размер которых варьируется от 7 до 70 миллиардов параметров. Llama 2-Chat оптимизирована для использования в диалоговых системах.

Предложенные модели превосходят чат-модели с открытым исходным кодом по большинству тестируемых показателей и могут стать подходящей заменой моделям с закрытым исходным кодом.

Сегодня попытки оценить клинические знания моделей, как правило, основаны на автоматизированных оценках по ограниченным критериям. Не существует стандарта для оценки прогнозов моделей и логических заключений в широком спектре задач. Для решения этой проблемы в работе [4] представлен MultiMedQA – тест, объединяющий шесть существующих наборов данных с открытыми ответами на вопросы, охватывающих профессиональные медицинские осмотры, исследования и запросы потребителей, и HealthSearchQA, новый набор бесплатных ответов на медицинские вопросы, поиск по которым осуществляется онлайн. Авторы предложили систему оценки типовых ответов человеком по нескольким параметрам, включая достоверность, точность, возможный вред и предвзятость.

В [5] предложена платформа доверенного медицинского ИИ для цифровизации процессов в различных областях. Основной построения платформы является использование метода RAG (Retrieval Augmented Generation – дополненная расширенная генерация) – использования большой языковой модели для формирования ответа на пользовательский вопрос на основе дополнительной информации из доверенного источника. Такой подход позволяет RAG-системе сформировать ответ, содержащий информацию, которая была недоступна при обучении и подстройке (FineTuning). RAG извлекает набор соответствующих документов для данного входного запроса или контекста, а затем использует эти документы в качестве дополнительных входных данных для нейронной языковой модели, которая генерирует текст на выходе. Цель RAG – повысить фактическую точность, разнообразие и информативность сгенерированного текста путем включения знаний из извлеченных документов.

В [6,7] работе проанализированы различные методы генерации с расширенным поиском (RAG) для поиска документов в корпусе знаний и их выборочного добавления ко входным данным LLM для генерации ответа. Авторами представлен фреймворк Self-BioRAG для обработки биомедицинских текстов. На основе трех наборов тестовых данных для ответов на медицинские вопросы, проведена серия экспериментов с Self-BioRAG, результаты которых демонстрируют значительный прирост производительности, достигая абсолютного улучшения в среднем на 7,2% по сравнению с современной моделью с 7B параметров. Аналогичным образом, Self-BioRAG превосходит RAG в среднем на 8% по баллам Rouge-1 при составлении более квалифицированных ответов на

два теста с ответами на вопросы в развернутой форме.

работе [8] рассмотрены проблемы применения RAG для построения СППВР в здравоохранении по следующим причинам:

- RAG полагается на качество и актуальность полученных документов, которые не всегда могут быть доступны или точны для конкретных областей или задач здравоохранения;
- RAG может быть не в состоянии охватить сложный и детализированный контекст медицинских сценариев, таких как предпочтения, ценности, цели, эмоции или социальные факторы пациента;
- RAG не может использоваться для обобщения истории болезни пациента, поскольку может оказаться невозможным извлечь наиболее актуальную и важную информацию из полученных документов, которые могут содержать много шума, избыточности или несоответствий;

Таким образом, RAG подходит не для всех приложений с генеративным ИИ в здравоохранении, и может потребовать тщательного проектирования, оценки и адаптации, чтобы обеспечить его безопасность, надежность и эффективность в конкретных условиях и задачах здравоохранения.

### 3. Разработка системы

Официальный ресурс Министерства здравоохранения Российской Федерации предоставляет доступ к обширной базе клинических рекомендаций, методических руководств и других медицинских документов, которые являются критически важными для обеспечения стандартизированной и качественной медицинской помощи в России.

Основные типы документов, доступные на сайте, включают:

- клинические рекомендации: эти документы содержат стандартизированные подходы к диагностике, лечению и профилактике различных заболеваний. Они служат руководством для врачей и медицинских учреждений, обеспечивая единообразие и высокие стандарты медицинской практики;
- методические руководства: включают подробные инструкции и методологии для выполнения медицинских процедур и исследований. Эти документы обеспечивают практическую поддержку медицинским работникам в их повседневной деятельности;
- справочные материалы: включают номенклатуры, классификаторы и другие

документы, которые помогают в организации и систематизации медицинской информации.

Документы на сайте доступны в формате PDF, что обеспечивает высокую степень сохранности оригинального форматирования и структурированной информации.

Задача состоит в том, чтобы найти ответ на заданный вопрос в тексте и оценить правильность найденного ответа, сравнивая его с ответом, предоставленным пользователем.

Дано:

- $T = \langle t_1, t_2 \dots t_n \rangle$  - текст, состоящий из предложений  $t_i$  ( $i = 1, 2, \dots, n$ );
- $Q$  - вопрос, на который необходимо найти ответ в тексте;
- $A_Q$  - вектор вложений, соответствующий сгенерированному ответу на вопрос  $Q$ ;
- $A_U$  - вектор вложений, соответствующий ответу пользователя на вопрос  $Q$ ;

- $\theta$  - порог сравнения.

Необходимо:

- найти ответ  $A_Q$  в тексте  $T$ , используя модель поиска и извлечения информации для определения текстового фрагмента, который наилучшим образом отвечает на вопрос  $Q$ ;
- сравнить ответ  $A_Q$ , найденный в тексте, с ответом пользователя  $A_U$  и оценить степень соответствия между ними.

На рисунке 1 приведена структура прототипа программного обеспечения, использующего модель векторизации и LLM (Large Language Model) для обработки вопросов пользователя и генерации ответов на основе семантически близкого контекста. Система включает следующие компоненты: большие языковые модели, механизм поиска релевантных фрагментов в текстовой базе, механизм генерации ответа.

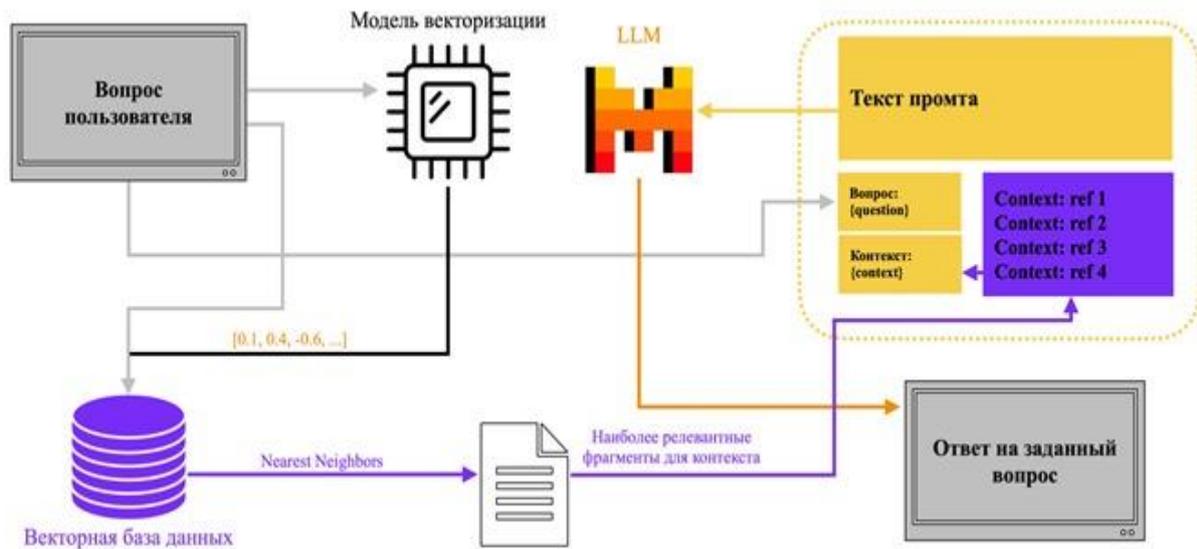


Рис. 1. Структура программы

Основные критерии для оценки правильности ответа:

1. релевантность найденного ответа по отношению к вопросу;
2. сходство найденного ответа с ответом пользователя.

Алгоритм анализа и генерации ответа включает следующие основные шаги:

1. ввод вопроса, подготовка предварительной базы документов;
2. разбивка базы документов на предложения;
3. оценка семантической близости каждого предложения в текстах к вопросу;

4. поиск наиболее схожего с вопросом по смыслу предложения;
5. формирование ответа на основе наиболее близких по смыслу предложений;
6. сравнение подготовленного моделью ответа с эталонным ответом пользователя;
7. оценка ответа пользователем.

Для оценки семантического сходства предложения  $t_i$  и вопроса  $Q$  (1) используется выражение:

$$R(t_i, Q) = \text{Relevance}(t_i, Q). \quad (1)$$

Для оценки сходства найденного ответа  $A_Q$  с ответом пользователя  $A_U$  (2) применяется следующее выражение:

$$S(A_Q, A_U) = \frac{A_Q * A_U - |GS_i - UGR_{iy}|}{\|A_Q\| \|A_U\|}. \quad (2)$$

Задание порогового значения для оценки правильности ответа модели (3):

$$\begin{cases} true, & \text{если } S(A_Q, A_U) \geq \theta \\ false, & \text{если } S(A_Q, A_U) < \theta \end{cases}. \quad (3)$$

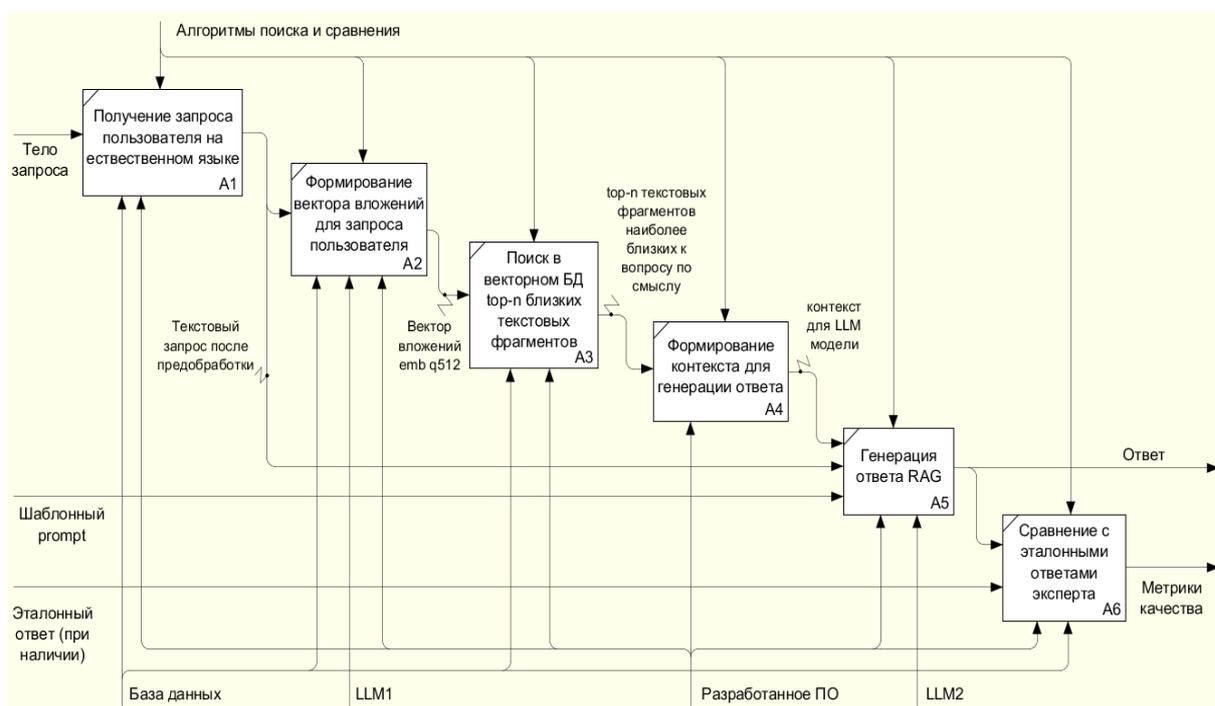


Рис. 2. Декомпозиция диаграммы IDEF0

На рис. 3 представлена общая схема базы данных, в которой хранятся тексты, вопросы, эталонные ответы.

Разрабатываемое приложение представляет собой комплексную систему для работы с текстами на основе технологий анализа естественного языка (NLP). Приложение включает следующий базовый функционал: загрузку и подготовку данных, создание и использование векторного хранилища фрагментов документов, генерацию ответов на запросы и оценку NLP метрик.

Сбор данных:

- источники данных: основными источниками данных являются клинические рекомендации и другие медицинские документы,

На рис. 2 представлена функциональная модель системы в нотации IDEF0.

Обработка данных выполняется с помощью следующих инструментов:

- LLM<sub>1</sub> – Multilingual-e5-large.
- LLM<sub>2</sub> – Mistral7B-qo8 + Taiga.
- Базы данных и разработанное ПО.

размещенные в формате PDF на сайте [sr.minzdrav.gov.ru](http://sr.minzdrav.gov.ru);

- методы сбора: для сбора данных используются ручные методы, включающие скачивание PDF-файлов непосредственно с сайта.

Преобразование данных:

- ручное преобразование документов PDF в текстовый формат TXT;
- очистка и нормализация текста: извлеченные тексты проходят этап предварительной обработки для удаления лишних символов, исправления ошибок и нормализации текста.

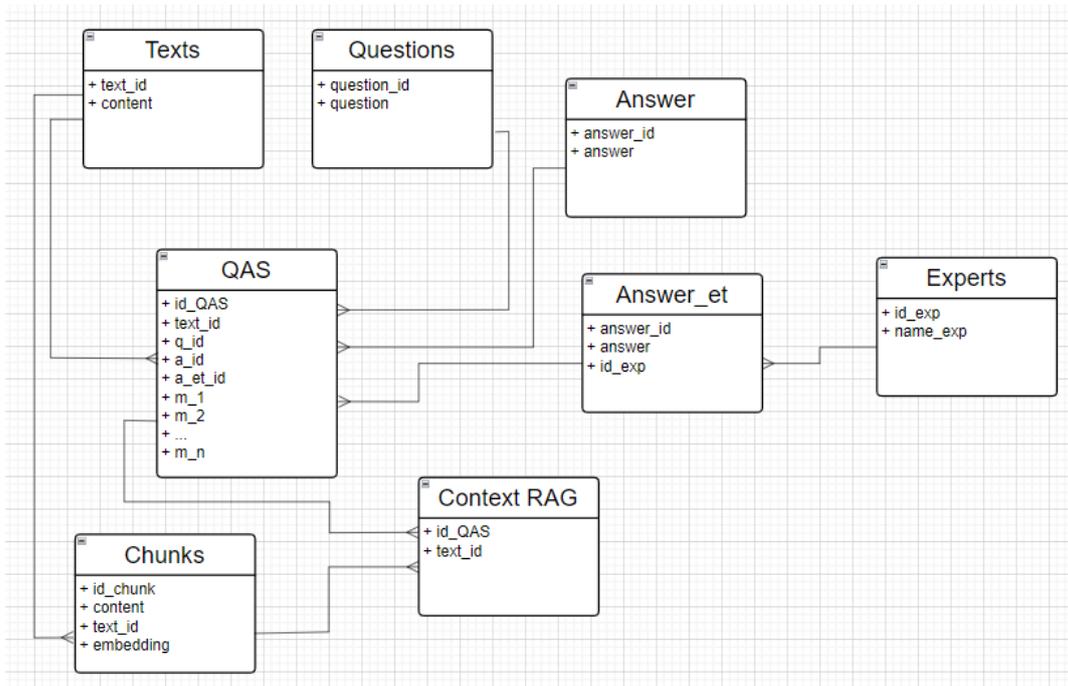


Рис. 3. Схема базы данных

#### 4. Вычислительный эксперимент

Генерация ответов на вопросы, заданные по текстовым документам, реализована с помощью библиотеки LangChain и большой языковой модели

Mistral. После генерации ответы записываются в файл, структура которого показана на рис. 4.

id	file_name	full_file_name	raw_text	text_num	file_name_q	full_file_name_q	raw_text_q	q_num	file_name	full_file_name	raw_text	AG_answer
0	1_T.txt	data\big_med\Te<блoк 1><блo	1 1_Q1.txt	1	data\big_med\QueЧто такое алл	1 1_A1.txt	1	1	data\big_med\AАллергический ринит (АР) – заболеван	Аллергический ринит - это воспалительное заболева		
1	1_T.txt	data\big_med\Te<блoк 1><блo	1 1_Q2.txt	1	data\big_med\QueФактор забол	2 1_A2.txt	2	1	data\big_med\AПыльца растений	Сенсибилизация к аллергенам домашней пыли, эпид		
2	1_T.txt	data\big_med\Te<блoк 1><блo	1 1_Q3.txt	1	data\big_med\QueОсновные пер	3 1_A3.txt	3	1	data\big_med\AВесенний (апрель-май) – связан с пылен	Весенний - апрель-май, ранний летний - июнь-серед		
3	1_T.txt	data\big_med\Te<блoк 1><блo	1 1_Q4.txt	1	data\big_med\QueКакие бывают	4 1_A4.txt	4	1	data\big_med\ALегкая степень – у пациента имеются сл	Аллергический ринит может быть легким, средним и		
4	1_T.txt	data\big_med\Te<блoк 1><блo	1 1_Q5.txt	1	data\big_med\QueКакие симпто	5 1_A5.txt	5	1	data\big_med\AZаложенность носа (обструкция), харак	Чихание, заложенность носа, водянистые выделени		
5	2_T.txt	data\big_med\Te<блoк 1><блo	2 2_Q1.txt	2	data\big_med\QueЧто такое бро	1 2_A1.txt	1	2	data\big_med\Aгетерогенным заболеванием, характери	Бронхиальная астма - это заболевание, при котором		
6	2_T.txt	data\big_med\Te<блoк 1><блo	2 2_Q2.txt	2	data\big_med\QueКакие характе	2 2_A2.txt	2	2	data\big_med\Aвсistingие хрипы, одышка, чувство зало	Свистящее дыхание, кашель, одышка, стеснение или		
7	2_T.txt	data\big_med\Te<блoк 1><блo	2 2_Q3.txt	2	data\big_med\QueЛабораторны	3 2_A3.txt	3	2	data\big_med\AРекомендуется проведение развернуто	Лабораторные диагностические исследования у взр		
8	2_T.txt	data\big_med\Te<блoк 1><блo	2 2_Q4.txt	2	data\big_med\QueСкорость разв	4 2_A4.txt	4	2	data\big_med\Aможет значительно варьировать у разных	Обострения БА могут развиваться как у пациентов с		
9	2_T.txt	data\big_med\Te<блoк 1><блo	2 2_Q5.txt	2	data\big_med\QueОрганизация	5 2_A5.txt	5	2	data\big_med\Aврач-аллерголог-иммунолог и/или врач-	Организация оказания медицинской помощи Диагн		
10	3_T.txt	data\big_med\Te<блoк 1><блo	3 3_Q1.txt	3	data\big_med\QueКак расшифро	1 3_A1.txt	1	3	data\big_med\AОстрые респираторные вирусные инфек	ОРВИ - это острое респираторное вирусное заболе		
11	3_T.txt	data\big_med\Te<блoк 1><блo	3 3_Q2.txt	3	data\big_med\QueИнкубационн	2 3_A2.txt	2	3	data\big_med\Aсоставляет от 1 до 14 суток	Инкубационный период при острых респираторных		
12	3_T.txt	data\big_med\Te<блoк 1><блo	3 3_Q3.txt	3	data\big_med\QueРекомендуетс	3 3_A3.txt	3	3	data\big_med\Aобщее недомогание, повышение темпе	росты состояния пациента перед лечением; угрожающие ж		
13	3_T.txt	data\big_med\Te<блoк 1><блo	3 3_Q4.txt	3	data\big_med\QueРекомендуетс	4 3_A4.txt	4	3	data\big_med\Aврача-отоларинголога (при наличии син	Нет, я не знаю ответа.		
14	3_T.txt	data\big_med\Te<блoк 1><блo	3 3_Q5.txt	3	data\big_med\QueЧто влияет на	5 3_A5.txt	5	3	data\big_med\Aпериод заболевания; тяжесть заболева	Выбор тактики лечения ОРВИ оказывается зависими		
15	4_T.txt	data\big_med\Te<блoк 1><блo	4 4_Q1.txt	4	data\big_med\QueЧто такое остр	1 4_A1.txt	1	4	data\big_med\Aострое инфекционное воспаление сли	Острый тонзиллофарингит - это инфекционное забо		
16	4_T.txt	data\big_med\Te<блoк 1><блo	4 4_Q2.txt	4	data\big_med\QueКак ставится д	2 4_A2.txt	2	4	data\big_med\Aжалоб больного на выраженную боль в	Инструментальные диагностические исследования.		
17	4_T.txt	data\big_med\Te<блoк 1><блo	4 4_Q3.txt	4	data\big_med\QueКакие основн	3 4_A3.txt	3	4	data\big_med\Aболь в горле, усиливающаяся при глот	ОТФ - это процесс, в ходе которого происходит перед		
18	4_T.txt	data\big_med\Te<блoк 1><блo	4 4_Q4.txt	4	data\big_med\QueКак проводит	4 4_A4.txt	4	4	data\big_med\AПри дифференциальной диагностике ви	Диагностика профессиональной БА. С целью диагно		
19	4_T.txt	data\big_med\Te<блoк 1><блo	4 4_Q5.txt	4	data\big_med\QueКакие ограни	5 4_A5.txt	5	4	data\big_med\Aострое, кислое, соленое, горячее, холо	д ограничения в еде при ОТФ зависят от индивидуаль		
20	5_T.txt	data\big_med\Te<блoк 1><блo	5 5_Q1.txt	5	data\big_med\QueЧто такое остр	1 5_A1.txt	1	5	data\big_med\Aвоспаление гортани и тканей подсклад	Острый обструктивный ларингит - это воспаление го		
21	5_T.txt	data\big_med\Te<блoк 1><блo	5 5_Q2.txt	5	data\big_med\QueЗначимыми в	2 5_A2.txt	2	5	data\big_med\Aреспираторные вирусы, причем до 80%	с сенсибилизаторы ведут себя как раздражители. Наибо		
22	5_T.txt	data\big_med\Te<блoк 1><блo	5 5_Q3.txt	5	data\big_med\QueПри осмотре	3 5_A3.txt	3	5	data\big_med\Aналичие признаков респираторного дис	со стороны врачей первичного звена и участковых те		
23	5_T.txt	data\big_med\Te<блoк 1><блo	5 5_Q4.txt	5	data\big_med\QueПри сборе ана	4 5_A4.txt	4	5	data\big_med\Aналичие атопии у ребенка, приступов бр	При сборе анамнеза на аллергический ринит нужно		
24	5_T.txt	data\big_med\Te<блoк 1><блo	5 5_Q5.txt	5	data\big_med\QueКакие профи	5 5_A5.txt	5	5	data\big_med\Aтщательное мытье рук после контакта с	1. Избегать переохлаждений. Особенно важно держ		
25	6_T.txt	data\big_med\Te<блoк 1><блo	6 6_Q1.txt	6	data\big_med\QueЧто такое остр	1 6_A1.txt	1	6	data\big_med\Aострое воспаление слизистой оболоч	Острый синусит - это воспаление слизистой оболоч		
26	6_T.txt	data\big_med\Te<блoк 1><блo	6 6_Q2.txt	6	data\big_med\QueКак клиничес	2 6_A2.txt	2	6	data\big_med\Aстойкими выделениями из носа, затруд	Острый синусит проявляется различными симптома		
27	6_T.txt	data\big_med\Te<блoк 1><блo	6 6_Q3.txt	6	data\big_med\QueКакие критери	3 6_A3.txt	3	6	data\big_med\AZаболевание с острым началом, сопров	Заболевание с острым началом, сопровождающееся		

Рис. 4. Фрагмент результатов обработки текстов в файле RAG

На рис. 5 представлен пример пользовательского ответа и ответа, сгенерированного программой.

raw_text	AG_answer
Аллергический ринит (АР) – заболевание, характеризующееся Ig	Аллергический ринит - это воспалительное заболевание слизистой носа, обусл
Пыльца растений	Сенсибилизация к аллергенам домашней пыли, эпидермальным аллергенам ил
Весенний (апрель-май) – связан с пылением деревьев (береза, с	Весенний - апрель-май, ранний летний - июнь-середина июля, поздний летний
Легкая степень – у пациента имеются слабовыраженные симпто	Аллергический ринит может быть легким, средним и тяжелым. Легкий ринит ха
Заложенность носа (обструкция), характерное дыхание ртом, сс	Чихание, заложенность носа, водянистые выделения из носа, зуд в полости но
гетерогенным заболеванием, характеризующимся хроническим	Бронхиальная астма - это заболевание, при котором в дыхательных путях прои
свистящие хрипы, одышка, чувство заложенности в груди и каш	Свистящее дыхание, кашель, одышка, стеснение или заложенность в груди.
Рекомендуется проведение развернутого общего (клинического	Лабораторные диагностические исследования у взрослых. Рекомендуется про
может значительно варьировать у разных пациентов – от неско	Обострения БА могут развиваться как у пациентов с уже известным диагнозом
врач-аллерголог-иммунолог и/или врач-пульмонолог, также – в	Организация оказания медицинской помощи Диагностику и ведение пациентс
Острые респираторные вирусные инфекции	ОРВИ - это острое респираторное вирусное заболевание.
составляет от 1 до 14 суток	Инкубационный период при острых респираторных вирусных инфекциях соста
общее недомогание, повышение температуры тела, утомляемо	состояние пациента перед лечением; угрожающие жизни острое состояние/за

Рис. 5. Пользовательский ответ и сгенерированный ответ

Далее сравниваются сгенерированные ответы и эталонные экспертные ответы по метрикам BERTScore, Saccor Blue, Meteor, Blue.

BERTScore – оценивает семантическую схожесть между сгенерированным и референсным текстами, используя вектора вложений, полученные с помощью предобученной модели трансформера BERT.

METEOR – метрика оценки качества машинного перевода, которая учитывает точные совпадения, синонимы и формы слов.

BLEU – метрика для оценки качества машинного перевода, основанная на подсчете совпадений n-

AG_answer	rouge1	rouge2	rougeL	rougeLsum	bertscore_pre	bertscore_rec	bertscore_f1	sacrebleu	meteor	bleu
Аллергиче	0	0	0	0	0,931755483	0,913231969	0,922400773	9,920262	0,337	0,103555
Сенсибил	0	0	0	0	0,78221494	0,856484056	0,817666411	0	0	0
Весенний	0	0	0	0	0,928645313	0,81816721	0,869912624	0,652771	0,101	0
Аллергиче	0	0	0	0	0,893078625	0,892552197	0,892815292	0,964516	0,071	0
Чихание,	0	0	0	0	0,901548266	0,796445489	0,845744014	0,040223	0,093	0,000525
Бронхиал	0	0	0	0	0,892992854	0,914920449	0,903823674	1,028497	0,163	0
Свистяще	0	0	0	0	0,939377487	0,937948823	0,938662648	15,13322	0,502	0
Лаборато	0	0	0	0	0,904723823	0,941749573	0,92286545	22,41815	0,666	0,224181

Рис. 6. Результат работы ПО в файле RAG\_metrics

Разработанная система продемонстрировала хорошие результаты по метрике BERTScore = 0,88, что подтверждает высокую семантическую схожесть между сгенерированными и эталонными ответами. По метрикам METEOR, BLEU и SacreBLEU результаты несколько ниже, что обусловлено различиями текстов на уровне слов и синтаксических конструкций. Однако данные метрики не являются абсолютными (см. табл. 1):

Таблица 1

Значения метрик

Метрики	Bertscore	Meteor	Blue	SacreBLEU	ROUGE
Среднее значение	0,88	0,31	0,11	12,86	-

грамм между сгенерированным и референсным текстами.

SacreBLEU – стандартизирует процесс вычисления метрики BLEU, чтобы обеспечить воспроизводимость результатов.

ROUGE – набор метрик, используемых для оценки качества автоматического обобщения текста.

После сравнения ответов по метрикам результаты записываются в файл, структура которого показана на рис. 6.

## 5. Заключение

Описаны подходы к построению системы поддержки принятия решений в медицинской практике, автоматизации процессов обработки данных. Предлагаемая структурно-функциональная организация вопрос-ответной системы для конкретной предметной области, основанная на методах машинного обучения, позволяет перейти к проектированию архитектуры прототипа программной системы.

Показана возможность применения больших языковых моделей при генерации ответов на пользовательские запросы по массиву текстовых документов с возможностью оценки их релевантности в составе систем поддержки принятия врачебных решений.

## Финансирование

Работа выполнена при финансовой поддержке РФФ (проект № 22-19-00471).

## Список используемых источников

1. Андрейченко А. Е., Гусев А. В. Перспективы применения больших языковых моделей в здравоохранении // Национальное здравоохранение. – 2024. – Т. 4. – №. 4. – С. 48–55.
2. Rohan Anil et al. PaLM 2 Technical Report. URL: <https://arxiv.org/abs/2305.10403>. (дата обращения 09.09.2024).
3. Hugo Touvron et al. Llama 2: Open Foundation and Fine-Tuned Chat Models. URL: <https://arxiv.org/abs/2307.09288/> (дата обращения 09.09.2024).
4. Karan Singhal et al. Large Language Models Encode Clinical Knowledge. URL: <https://arxiv.org/abs/2212.13138> (дата обращения 09.09.2024).
5. Платформа доверенного ИИ в медицине. URL: <https://healthops.ru/> (дата обращения 06.09.2024).
6. Minbyul Jeong, Jiwoong Sohn, Mujeen Sung, Jaewoo Kang, Improving medical reasoning through retrieval and self-reflection with retrieval-augmented large language models, *Bioinformatics*, Volume 40, Issue Supplement\_1, July 2024, Pages i119–i129, URL: <https://doi.org/10.1093/bioinformatics/btae238> (дата обращения 09.09.2024).
7. Jeong M. et al. Improving medical reasoning through retrieval and self-reflection with retrieval-augmented large language models. URL: <https://arxiv.org/abs/2401.15269>. – 2024 (дата обращения 09.09.2024).
8. Eapen BR. (May 11, 2024). Nuchange.ca - Why is RAG not suitable for all Generative AI applications in healthcare? URL: <https://nuchange.ca/2024/05/why-is-rag-not-suitable-for-all-generative-ai-applications-in-healthcare.html>. (дата обращения 09.09.2024).