RU EN ПРОГРАММНЫЕ СИСТЕМЫ: ТЕОРИЯ И ПРИЛОЖЕНИЯ ISSN 2079-3316
научная статья ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ В МЕДИЦИНЕ

UDC 61:007.51:316
10.25209/2079-3316-2025-16-6-197-219

Автоматическое распознавание речевых медицинских данных с использованием LLM

Юрий Геннадьевич **Сидоров**[1], Владимир Леонидович **Малых**[2], Алексей Николаевич **Калинин**[3], Ольга Сергеевна **Елистратова**[4]

[1,3] Группа компаний «Интерин», Москва, Россия
[2,4] Институт программных систем им. А. К. Айламазяна РАН, Веськово, Россия

Аннотация. Одним из барьеров, препятствующих широкому распространению голосового ввода медицинских данных в МИС, являются недостаточные потребительские качества текстов, получающихся после транскрибации. Не все медицинские термины и слова общего лексикона распознаются корректно, нарушается согласование слов по роду, числу и падежам, текст недостаточно хорошо форматирован с точки зрения грамматики. Всё это требует дальнейшей доработки текста. Ещё одной сложной проблемой видится необходимость приведения текста к структуре медицинского документа в МИС. Структура документа может быть достаточно сложной, содержать много элементов, иметь требования к типу и формату данных в элементах структуры. Речевой ввод может лишь частично использоваться для формирования документа, а недостающие данные могут быть взяты из пользовательского шаблона.

Для решения указанных проблем предлагается использовать LLM в качестве корректора результатов транскрибации речи, интегратора речевых и текстовых данных из шаблона и формирователя структуры результирующих данных. В работе предлагается архитектура решения для ввода речевых медицинских данных на основе композиции системы транскрибации и LLM. Предлагается методика проведения испытаний решения, включающая подготовку набора данных и метрику расчёта качества решения. Описывается реализация решения на основе свободной и проприетарной компонент.

Результаты могут быть использованы при разработке и оценке систем ИИ, применяемых для ввода речевых данных, и не только в медицине.

Ключевые слова и фразы: медицинские информационные системы, МИС, искусственный интеллект, ИИ, речевой ввод, система транскрибации, большие языковые модели, LLM

Благодарности: Авторы искренне благодарны Дмитрию Владимировичу Белышеву за внимание к статье и ценные замечания, позволившие улучшить содержание статьи

Для цитирования: Сидоров Ю.Г., Малых В.Л., Калинин А.Н., Елистратова О.С. *Автоматическое распознавание речевых медицинских данных с использованием LLM* // Программные системы: теория и приложения. 2025. **Т. 16**. №6(71). С. 197–219. https://psta.psiras.ru/read/psta2025_6_197-219.pdf

Введение

Современные информационные технологии совершают на наших глазах очередную научно-техническую революцию в области искусственного интеллекта. Во все сферы жизни проникают интеллектуальные приложения, пользователям предлагаются интеллектуальные интерфейсы для информационного взаимодействия. Никого уже не удивляют возможности голосового управления различными устройствами, голосовой ввод запросов в ПК и мобильные телефоны, исполнение автомобилем голосовых команд, речевое общение с интеллектуальными ботами и т. п. Речь, голосовой ввод данных стали широко и повсеместно использоваться.

Эта тенденция затронула также и сферу медицины. Однако в сфере медицины имеются свои особые требования к возможностям и качеству работы интеллектуальных речевых интерфейсов. В ранней работе [1] указывалось на возможности и перспективы использования речевых интерфейсов в МИС. Ещё 10 лет назад эти возможности не были востребованы в отечественной медицинской информатике, запись на диктофон с последующей обработкой данных человеком мы не относим к поставленной проблеме. По мере совершенствования программ распознавания речи в сфере медицинской информатики появляется всё больше решений для автоматизации речевого ввода медицинских данных в МИС. По данным Яндекса (поиск с помощью Яндекс GPT):

На данный момент в российских медицинских учреждениях используются следующие информационные системы с поддержкой речевого ввода:

- *ЕМИАС* (Единая медицинская информационно-аналитическая система) Москвы. Интегрирована с решением Voice2Med, используется более чем в 120 медицинских учреждениях столицы. Система обрабатывает ежемесячно свыше 212 тысяч минут речи врачей.
- *МИС qMS* в интеграции с *Voice2Med*. Применяется в Национальном медицинском исследовательском центре имени В. А. Алмазова. Позволяет работать в режиме «одного окна» без использования сторонних приложений.
- *ЕМИАС Московской области*. Интегрирована с системой голосового ввода от ООО «ЦРТ». В системе работает более 2100 гарнитур голосового ввода, что позволило увеличить эффективность внесения информации на 47%.

Технология *Voice2Med* является основным решением для речевого ввода в российских МИС. Её ключевые особенности:

- Распознавание русской речи с точностью до 97–98%.

- Поддержка специализированных медицинских терминов и сокращений.
- Возможность работы как со стационарными компьютерами, так и с планшетами.
- Интеграция с различными медицинскими информационными системами.

С одной стороны, можно говорить о значительных успехах в области разработки решений для речевого ввода медицинских данных на русском языке. С другой стороны, обращение к зарубежному опыту, а также проведённые авторами эксперименты по речевому вводу медицинских данных позволяют говорить о нерешённых проблемах в данной сфере. В источнике [2] отмечается, что голосовой помощник в смартфоне справляется с повседневными разговорами с точностью 95%, но в больнице точность падает до 70–80 %.

Основная проблема в профессиональном языке, на котором говорят врачи. Традиционные модели преобразования речи в текст обучаются на огромных массивах данных, собранных из интернета, аудиокниг и повседневных разговоров. Медицинская терминология почти не встречается в этих обучающих данных. Медицинские термины не просто звучат по-другому – они подчиняются совершенно иным лингвистическим правилам. В названиях фармацевтических препаратов латинские корни сочетаются с современной химией. Анатомические термины состоят из нескольких слогов и требуют точного произношения. А медицинские аббревиатуры – это минное поле.

В клинических условиях всё становится ещё хуже. В отделениях неотложной помощи срочные разговоры заглушаются сигналами оборудования. В операционных несколько говорящих в масках. Консультации в отделении интенсивной терапии проходят под шум аппаратов искусственной вентиляции лёгких. Стандартное автоматическое распознавание речи предполагает чистый звук с чётким разделением говорящих, а не контролируемый хаос, характерный для реального здравоохранения.

Аналогичные проблемы речевого ввода отмечаются и в работе [3]. «Ошибки в системах автоматического распознавания речи в медицине разнообразны и проблематичны: от неправильного толкования названий лекарств и дозировок до неверных результатов лабораторных исследований, анатомических ошибок, несоответствий возраста и пола и даже неправильных имен врачей и формата дат.

Дополнительные проблемы включают в себя генерацию бессмысленных слов, а также пропуски и дублирование. Эти неточности могут иметь

серьёзные последствия, потенциально ставя под угрозу диагностику пациентов и принимаемые решения о лечении. Для преодоления этих ограничений требуются инновационные решения, выходящие за рамки текущих возможностей традиционных систем автоматического распознавания речи».

Можно только согласиться с этими утверждениями, так как собственные эксперименты авторов с реальными клиническими данными всё это подтвердили. Даже в идеальных условиях ввода данных при отсутствии посторонних шумов и посторонней речи мы столкнулись с проблемой распознавания медицинских терминов.

Далее будут приведены некоторые примеры ошибок транскрибации в медицинских текстах. Можно было бы привести полный список ошибок систем транскрибации полученных авторами на отобранных для экспериментов реальных клинических данных, но оценка качества распознавания различных систем не является целью статьи. Можно лишь констатировать наличие проблем с медицинскими речевыми данными.

В [2] критически упомянуты пути к улучшению распознавания речевых медицинских текстов:

- Обучение на пользовательском словаре. Требует специализированных наборов данных и постоянного обновления по мере развития медицинских знаний.
- Системы исправления ошибок после обработки на основе правил поверх ошибочных транскрипций. Часто приводит к появлению новых ошибок.
- Специализированные медицинские модели. Стоят дорого и ограничены узкими сценариями использования, имеют проблемы с обобщением и контекстным пониманием.
- Методы повышения релевантности слов кажутся эффективными для улучшения распознавания конкретных терминов, но использование очень длинных списков слов противоречит изначальной цели повышения релевантности конкретных слов (то есть 98% слов будут отвлекающими факторами).

Есть и другой обнадёживающий путь к повышению качества речевого ввода медицинских данных. Этот путь связан с применением больших языковых моделей – LLM (Large Language Model). Широкие возможности LLM впечатляют: информативные ответы на вопросы, перевод с одного языка на другой, обобщение и аннотация больших текстов, для медицины поддержка принятия врачебных решений, постановка диагнозов и многое другое. Применительно к речевому вводу, недавно одному из авторов,

участвующих в голосовой конференции, по её окончанию было предложено сформировать обобщенное представление состоявшейся беседы.

Имеются зарубежные и отечественные обзоры, рассматривающие применение LLM в сфере медицины [4,5]. В [4] выделены различные сценарии применения LLM. Забегая вперёд, отметим, что теме и содержанию статьи соответствует сценарий «Electronic health record, clinical letters and medical note generation», в русскоязычном изложении «Электронная медкарта, формирование клинических медицинских записей».

LLM безусловно являются технологическим прорывом, затрагивающим и сферу медицины. В работе [6], посвящённой проблемам поддержки принятия врачебных решений, отмечалась настоятельная необходимость в формировании общего глобального подхода к решению проблемы, подхода, не опирающегося на частные маргинальные решения.

Можно считать, что с появлением LLM такой подход появился. Надежды на ИИ и, в частности, LLM настолько велики, что США представили план глобального доминирования в сфере ИИ [7]. Основная цель – «неоспоримое и неоспариваемое технологическое доминирование» США в сфере ИИ. Эксперты отмечают, что документ впервые официально закрепляет стратегию «техноимпериализма» – создания глобальной системы зависимости от американских технологий.

Использование LLM в отечественной медицинской информатике – это ближайшее будущее. Вместе с тем, следует отметить определённую настороженность по отношению к LLM в сфере медицины. Многие эксперты настроены критически по отношению к LLM, это связано с присущей LLM способностью «галлюцинировать», т. е. просто выдумывать или фантазировать [4,5].

Как отмечают специалисты, в 3%–5% случаев LLM может выдавать на ваши вопросы совершенно неадекватные ответы, что совершенно не допустимо для сферы здравоохранения. Это утверждение верно, когда мы ставим вопрос о принятии врачебного решения на основании «мнения» LLM (постановка диагноза, выбор лечебно-диагностических мероприятий). Но мы можем использовать LLM не для принятия решений, а как технический инструмент, ошибки которого легко исправимы и не критичны.

В проблеме речевого ввода медицинских данных напрашивается очевидное решение – использовать LLM в качестве корректора данных транскрибации. В [2, 3] и [8], собственно, это и предлагается. Сила такого решения в том, что LLM при коррекции текста работает не по отдельности с каждым словом, но использует контекст всего текста.

Контекстуальное понимание LLM позволяет элиминировать неверно распознанные слова и даже позволяет «восстановить» правильные потерянные при транскрибации слова.

В наших вычислительных экспериментах мы столкнулись с тем, что система транскрибации «не поняла» в речи термин «юкставезикального», который заменила на «щитовидной кального». Результат транскрибации был отправлен в LLM DeepSeek с заданием исправить ошибки транскрибации. DeepSeek, используя контекст, смог элиминировать «щитовидной кального» и «восстановил» правильный термин «юкставезикального», при том, что этого слова не было во входных данных LLM, модель «знала» только то, что она работает с протоколом УЗИ мочевого пузыря. Такое поведение LLM действительно можно соотнести с интеллектуальным поведением человека.

Мы хотим поставить LLM более широкую задачу, чем коррекция текста после транскрибации. Одним из барьеров, препятствующих широкому распространению голосового ввода медицинских данных в МИС, являются недостаточные потребительские качества текстов, получающихся после транскрибации. Не все медицинские термины и слова общего лексикона распознаются корректно, нарушается согласование слов по роду, числу и падежам, текст недостаточно хорошо форматирован с точки зрения грамматики. Всё это требует дальнейшей доработки текста.

Ещё одной сложной проблемой видится необходимость приведения текста к структуре медицинского документа в МИС. Структура документа (осмотра, диагностического протокола) может быть достаточно сложной, содержать много элементов, иметь требования к типу и формату данных в элементах структуры. Обе проблемы видятся достаточно трудноразрешимыми при традиционном подходе к ним - обучение и совершенствование систем транскрибации, написание частных парсеров текстовых данных в структурированные шаблоны документов МИС.

Оказалось, что можно предложить и другой путь решения проблем путём обращения к LLM. Можно взять неструктурированный, полученный после транскрибации медицинский текст без форматирования, с ошибками транскрибации и передать его на обработку LLM, предоставив модели запрос, контекст и требуемую структуру результирующего документа. При этом необходимая структура документа хранится в МИС и может быть получена непосредственно из МИС, например в формате JSON.

Ещё одним важным требованием при формировании медицинских документов является использование шаблонов заполнения медицинских документов. Врачи охотно создают шаблоны для различных типовых медицинских случаев для предварительного заполнения документов.

Шаблоны сильно сокращают время и труд по подготовке документов, врачам остаётся внести в документ нешаблонные данные. Применительно к речевому вводу это означает, что документ будет формироваться не только на основании обработанных речевых данных, но и на основании данных шаблона заполнения.

В общем случае LLM должна выступать в роли корректора данных транскрибации, интегратора обработанных речевых и текстовых данных из шаблона, и как формирователь структуры данных (приведение текста к требуемой структуре). Столь широкая постановка функциональной задачи речевого ввода медицинских данных является новой и перспективной.

Дополнительным важным требованием к постановке задачи является требование использования при решении отечественных программных компонентов и свободно распространяемых компонентов с открытым кодом. Только так можно будет избежать пресса «техноимпериализма» [7].

Важную роль также играют ресурсные ограничения медицинских организаций. Следует рассмотреть достаточно «лёгкие» решения с точки зрения использования компьютерного оборудования. LLM с сотнями миллиардов параметров требуют для развёртывания достаточно большие вычислительные мощности, которые ни одна МО себе позволить не сможет. Следует рассмотреть применение лёгких дистиллированных LLM, специально обученных медицинских LLM. Можно рассматривать использование облачных решений, но не все МО могут их легко применять в силу специфики их закрытости от внешнего мира, примеры см. [9].

Следует учитывать и финансовые затраты МО на владение предлагаемой технологией. Подытожив сказанное, можем сформулировать цель работы.

Цель работы: разработка и исследование архитектуры обработки речевых медицинских данных с применением LLM в качестве корректора, интегратора и формирователя структуры данных. Разработка методологии проведения исследования, включая формирование набора данных и оценку качества исследуемого решения.

1. Шаблоны документов

Тема шаблонов документов требует отдельного рассмотрения, так как она важна для постановки задачи и проведения исследования. Можно выделить в этой теме два аспекта.

Шаблон структуры документа. Как правило, медицинские документы имеют структуру. Эта структура видна в печатных копиях

документов, структура отражается в объектах хранения и представления документов в базе данных МИС. Обычно структура представляет собой дерево. Информационной моделью структуры документа может служить xml или Json объект. На рисунке 1 представлена структура одного из документов, полученная путём обхода дерева в глубину.

Правая почка:
Контуры:
Толщина паренхимы:
Структура паренхимы почки:
В почечном синусе:
Чашечно-лоханочная система:
Левая почка:
Контуры:
Толщина паренхимы:
Структура паренхимы почки:
В почечном синусе:
Чашечно-лоханочная система:
Проекция надпочечников:
УЗ-ангиография:
Заключение:
Рекомендации:
Патологии:
Диагноз:

Рисунок 1. Шаблон структуры документа, переданный в LLM

Обратное отображение этой структуры в дерево тривиально при условии, что все узлы структуры имеют уникальные имена. Именно такие шаблоны структуры документа использовались нами при проведении исследования.

Шаблон заполнения документа. Такие шаблоны используются врачами повсеместно. Шаблон заполнения может быть совмещён со структурой документа - атрибуты узлов xml или пары ключ – значение Json. Основная сложность применения шаблонов заполнения как дополнительного к речевому источнику данных для LLM заключается в сложности «умного» синтеза данных из двух источников. Врачи при работе с данными из шаблона заполнения могут действовать по-разному с каждым из элементов данных:

(insert) оставлять неизменным;
(delete) удалять данные заполнения;

(append) объединять данные заполнения с дополнительными данными вводимыми пользователями;

(edit) отредактировать данные заполнения.

Очень легко реализовать общие стратегии обработки шаблона заполнения документа, применяемые ко всем полям документа:

(1) заполнить из шаблона поля, которые после обработки LLM остались пустыми (insert if empty);

(2) добавить данные из пользовательского шаблона ко всем полям документа после обработки LLM (append).

В этом случае нет нужды в передаче LLM шаблона заполнения документа. Обработка шаблона делается после получения от LLM текста структурированного документа. Гораздо сложнее «добиться» от LLM умного синтеза данных из речевого ввода с данными из шаблона заполнения, синтеза эмулирующего работу врача. Эта проблема нами изучается, исследования в этом направлении будут продолжены. В настоящей работе мы не использовали шаблоны заполнения документов, использовались только шаблоны структуры документов.

2. Подготовка набора данных для проведения исследования

Согласно [**10**] наибольшую ценность для тестирования решения представляют верифицированные врачами наборы данных (НД). Документы для проведённого исследования были подготовлены врачами и по содержанию и качеству они полностью соответствуют реальным клиническим документам. Поэтому можно утверждать, что исследование и оценка решения проведены на реальных клинических данных. В МИС семейства Интерин накоплены миллионы (сотни тысяч в рамках отдельной МО) клинических документов. Необходимы критерии для их отбора в НД. Мы предлагаем использовать следующие два критерия отбора.

Первый критерий: отбор документов наиболее часто используемых типов. В реестре документов МИС вычисляется частота документов различного типа. Далее выбирается некоторое множество часто используемых типов и выбирается число экземпляров каждого типа. Отобранные документы формируют НД.

Второй критерий: отбор документов по покрытию ими словаря. Отбираются документы (за определённый временной период, определённых типов, см. раздел 1. Тексты отобранных документов индексируются для создания словаря.

Далее задаётся число документов в НД. Первым выбирается документ, который содержит наибольшее количество слов из словаря, последующие

документы выбираются по правилу, что каждый следующий новый документ, включаемый в НД, максимально расширяет словарь (множество слов) в уже отобранном множестве документов. Таким образом можно получить наиболее полное покрытие лексикона отобранных документов – кандидатов в НД.

Было решено использовать подготовленные врачами для исследования медицинские клинические документы следующих типов: протоколы осмотров и диагностических исследований нескольких специалистов (ЛОР, Уролог, УЗИ-диагност). Такие данные, как было уже отмечено, можно отнести к верифицированному набору данных [**10**]. Подготовленные медицинские документы в количестве десяти различных экземпляров изначально не содержали персональных данных. Можно было формально считать, что эти документы полностью удовлетворяют требованиям ГОСТа [**10**] по деидентификации (de-identification, процесс удаления связи между совокупностью идентифицирующих данных и субъектом данных).

Чтобы уменьшить зависимость и корреляцию результатов от одного спикера (особенности речи и произношения), были приглашены пять дикторов (не врачи) для озвучки отобранных текстов в разных условиях записи аудио файлов (микрофоны различного качества, возможные посторонние шумы). В результате были записаны 50 аудио файлов. Для каждого экземпляра документа, участвующего в исследовании, был сформирован специализированный запрос (промт), контекст запроса и передаваемая LLM структура документа.

Отобранные документы и шаблоны структуры документов сформировали НД для исследования.

Отметим, что формирование НД вполне соответствовало первому критерию отбора документов. В МО протоколы УЗИ находились на третьей позиции по частоте использования (917505 экземпляров), осмотры отоларинголога находились на 9-й позиции (328110 экземпляров), осмотры уролога находились на 12-й позиции (257419 экземпляров).

3. Архитектура ввода речевых медицинских данных

Для концептуальной проверки подхода, базирующегося на LLM корректоре для устранения ошибок транскрибации и генерации структурированного документа по конкретному шаблону структуры, были разработаны и размещены на экспериментальном стенде модули взаимодействия с ASR (Automatic Speech Recognition) и LLM корректора (post-ASR), см. рисунок 2 . Модуль взаимодействия с ASR позволял обращаться к различным сервисам транскрибации: приватным и open source. Модуль

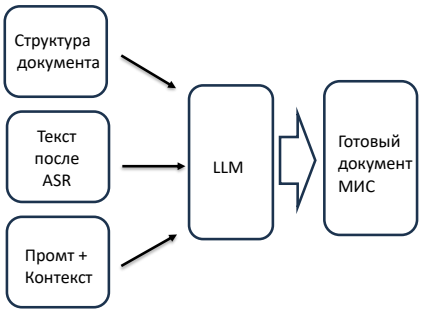


Рисунок 2. Входные данные LLM

post-ASR позволял обращаться к различным LLM: проприетарным в облаке и локально развёрнутыми open source LLM. Предложенную и реализованную на испытательном стенде архитектуру ввода речевых медицинских данных, соответствующую представленным на рисунках 2 и 3 схемам, будем называть далее решением.

Рисунок 3. Общая схема решения

В качестве кандидатов на использование в решении были выбраны: для транскрибации файлов локальная open source модель Whisper и облачные проприетарные системы Яндекса и Сбербанка, Speech2Text, для LLM корректора open source модели Gemma3 - 12b и Qwen3 - 8b, и облачные проприетарные системы Яндекс GPT, DeepSeek и др.

Было решено провести предварительное испытание двух вариантов решения на компонентах *транскрибация + LLM*:

(*1*) проприетарная облачная нейросеть Сбербанка SaluteSpeech + проприетарная облачная модель Yandex GPT;

(*2*) локальная модель Whisper Large от OpenAI + проприетарная облачная модель Yandex GPT (работа с моделью настроена через API с использованием библиотеки yandex_cloud_ml_sdk).

4. Методика оценки качества решения

Для оценки качества транскрибации аудио файлов существуют специальные метрики CER – частота ошибок в символах и WER – частота ошибок в словах. Для оценки качества решения эти метрики не могут быть использованы, по трём причинам:

(*1*) в предложенной постановке не все данные в результате формируются через голосовой ввод, часть данных может находиться в шаблонах;

(*2*) LLM может выдавать результат в несколько отличных, но семантически верных терминах;

(*3*) структуризация приводит к изменению порядка слов в тексте. Например, мы в ходе вычислительных экспериментов увидели замену исходного слова «двугорбой» на «двуглавой», замену в протоколе УЗИ мочевого пузыря единиц измерения объёма мочи с «мл» на «см. куб.». Все эти примеры замен корректны, но они будут считаться «искажением» оригинальной речи и будут вносить вклад в метрики CER и WER.

Поэтому предлагается ввести в рассмотрение другую метрику оценки качества.

В нашей постановке важна семантическая близость исходных данных (диктуемый диктором текст + шаблон структуры данных) к результату, полученному подачей на вход решения исходных данных.

Предлагается в исходном файле для речевого ввода выделить разметкой отдельные семантические единицы (объекты). Далее при оценке качества результата проверяется, присутствует ли в результате выделенная семантическая единица (СЕ). СЕ может присутствовать, может быть потеряна. Кроме того, LLM может сама нафантазировать СЕ, которых не было в изначальных данных. Следует учесть возможное зашумление исходных данных паразитными СЕ, имитирующими или отражающими постороннюю речь или технические реплики врача, которые не должны будут попасть в итоговый медицинский документ.

Рисунок 4 даёт пример разметки небольшого диагностического протокола УЗИ мочевого пузыря. Семантические единицы выделяются прямо в тексте маркером (1).

Соответственно, мы ожидаем что в результате обработки решением приведенного размеченного текста мы получим новый структурированный текст, содержащий все указанные СЕ и более ничего лишнего. Имеется определённая свобода в выборе СЕ при разметке. Можно за СЕ принимать целые высказывания (фразы, предложения), можно делить высказывания на отдельные, сохраняющие семантику, единицы.

Мочевой пузырь заполнен достаточно (1), содержит 240 мл мочи. (1)
Контуры ровные (1).
Содержимое гомогенное (1).
Стенки не утолщены (1).
Внутрипросветные образования не выявлены (1).
Мочеточниковые выбросы в режиме ЦДК регистрируются с обеих сторон (1).
Итого 7 семантических единиц

Рисунок 4. Размеченный протокол УЗИ мочевого пузыря

Выбор «гранулированности» СЕ повлияет на значение метрики. В целом семантика отдельного высказывания может быть не передана, но при этом сохранена семантика отдельных частей высказывания. Оставляем вопрос выбора СЕ на читателя, авторы предпочли в основном использовать элементарные СЕ, которые уже невозможно разбить на ещё более мелкие СЕ.

Шаблоны структуры документа не нуждались в семантической разметке, так как требование к LLM привести данные к указанной структуре было императивным и LLM это требование выполняла.

Подходящая для наших целей метрика представлена в работе [11]. Согласно [11] в каждом результате следует оценить (подсчитать) следующие критерии:

Истинно-положительные (true positives, tp) – ожидаемые результаты.

Ложно-положительные (false positives, fp) – ошибочные результаты в выдаче.

Ложно-отрицательные (false negatives, fn) – ожидаемые результаты, но не попавшие в выдачу.

Истинно-отрицательные (true negatives, tn) – результаты, которые не попали и не должны были попасть в выдачу.

Точность (P, precision), которая указывает на то, сколько точных результатов получено в выдаче, определяется по формуле:

$$P = \frac{tp}{tp + fp}$$

Полнота выдачи (R, recall):

$$R = \frac{tp}{tp + fn}$$

В качестве семантической метрики рекомендуется использовать унифицированную метрику F:

$$F = \frac{2 \cdot P \cdot R}{P + R}$$

4.1. Методика расчёта метрики:

(1) В исходном тексте мысленно выделяются и размечаются семантические единицы, которые должны присутствовать в результате. Например: «контуры ровные», «объем 280 мл», «подвижная структура гиперэхогенная 35 на 20, дающая акустическую тень».

(2) Далее определяется присутствие этих семантических единиц в результирующем тексте. Например: «Контуры: норма – чёткие, ровные», «Объём: 280 куб. см», «Объёмные образования: патология – конкремент d 35 на 20 мм, дающий акустическую тень». Если семантическая единица присутствует в результате, то $tp := tp + 1$. Если не присутствует, то $fn := fn + 1$.

(3) Далее выделяются семантические единицы, которые присутствуют в исходном тексте, но не должны присутствовать в результате. Например: «так ... повернитесь, секундочку». Если таковые семантические единицы не присутствуют в результате, то $tn := tn + 1$. Если таковые семантические единицы присутствуют в результате, то $fp := fp + 1$.

(4) LLM может нафантазировать какие-то семантические единицы, которых не было в исходном тексте и не было в шаблоне структуры документа. Для всех таких единиц $fp := fp + 1$.

Важно отметить, что при расчёте метрики выделяются только семантические ошибки. Грамматические ошибки, если они не приводят к потере семантики, в расчёт не берутся. Вот примеры из проведенных вычислительных экспериментов. «Анализ мочи цито» на выходе превратился в «Анализ мочи пациента», так как система транскрибации перевела «цито» в «пациента», а LLM это пропустила. Мы посчитали это семантической ошибкой, так как был потерян важный признак срочности проведения анализа. И другой пример. DeepSeek в осмотре урологом мужчины в результате выдал «яичко» в женском роде, «правая яичка», «левая яичка». Мы решили, что семантика в данном случае сохранена и это не семантическая ошибка. Конечно, такие грамматические ошибки подлежат исправлению и снижают качество результата, но на передачу семантики они не влияют.

5. Результаты исследования

Для работы с локальными моделями LLM был подготовлен стенд на локальном PC Intel(R) Core(TM) i7-6850K CPU 3.60GHz ОЗУ 16 Гб Графический процессор NVIDIA GeForce RTX 2070 SUPER 16,0 ГБ. Операционная система (OS) Windows 10, Python, библиотеки tensorflow, torch, ffmpeg-python, CUDA (для работы с GPU), yandex-cloud-ml, приложение для PC от Сбера SaluteSpeech App.

Были проведены вычислительные эксперименты с двумя решениями, которые символически названы Sber+YaGPT и Whisper+YaGPT. В решениях были использованы следующие компоненты:

Sber – нейросеть SaluteSpeech для распознавания и синтеза речи созданная Сбербанком.

YaGPT – нейросеть семейства GPT от компании «Яндекс».

Whisper – нейросеть, разработанная компанией OpenAI для автоматической расшифровки аудиозаписей и преобразования речи в текст.

Каждое решение обработало один и тот же набор данных из 50 различных аудио файлов (10 документов, зачитанных 5-ю дикторами). Набор данных включал в себя 1345 размеченных семантических единиц. Была проведена статистическая обработка результатов согласно предложенной методике оценки качества решения, результаты сведены в таблицу.

Таблица 1. Сравнительная статистика по решениям

Решение	Распознано СЕ, %		Потеряно СЕ, %		Метрика
Sber + YaGPT	1287	95,69	58	4,31	0,98
Whisper + YaGPT	1180	87,73	165	12,27	0,95

Таблица 2. Сравнительная статистика по дикторам на решении Sber+YaGPT

Диктор	Распознано СЕ, %		Потеряно СЕ, %		Метрика
D1	257	95,54	12	4,46	0,98
D2	257	95,54	12	4,46	0,98
D3	251	93,31	18	6,69	0,96
D4	261	97,03	8	2,97	0,99
D5	261	97,03	8	2,97	0,99

Дикторы оказывают влияние на точность распознавания СЕ. Максимальная разница в точности распознавания согласно данным таблицы 3 на решении Sber+YaGPT составила 3,72% или 10 СЕ в абсолютных единицах. Следует учесть, что все аудиофайлы записывались в благоприятных условиях без помех и шумов. В реальных условиях записи на рабочих местах врачей можно ожидать ухудшение точности распознавания СЕ.

Рисунки 1, 5, 6 и 7 дают пример обработки речевого ввода диагностического исследования почек решением Sber+YaGPT.

ОСМОТР ПОЧЕК:
Правая почка расположена обычно (1), размерами 9,9x3,8x3,8 см. (1)
Контуры ровные, чёткие. (1) Толщина паренхимы 1,7 см. (1)
Структура паренхимы почки несколько неоднородная. (1) Эхогенность обычная (1).
В почечном синусе - мелкие линейные гиперэхогенные структуры (1) без четких акустических теней (1) (уплотненные стенки сосудов).
Чашечно-лоханочная система не дилятирована. (1)
Левая почка несколько опущена (1), размерами 10,9x5,0x4,5 см (1). Контуры ровные, чёткие. (1)
Толщина паренхимы 1,8 см. (1)
Структура паренхимы почки несколько неоднородная (1). Эхогенность обычная (1). В сосудистых структурах в паренхиме киста (1) размерами 1,7x1,3x1,4 см (1).
В почечном синусе - мелкие линейные гиперэхогенные структуры (1) без четких акустических теней (1) (уплотненные стенки сосудов).
Чашечно-лоханочная система не дилятирована (1).
В проекции надпочечников патологические образования не определяются. (1)
При УЗ-ангиографии ход основных сосудистых структур почек обычный. (1)
В обеих почках во всех сегментах кровоток прослеживается до периферических отделов коркового слоя. (1)
Заключение: Киста левой почки. (1)
Рекомендации: Контроль УЗИ. (1)
Патологии:
Диагноз: Киста почки впервые выявлено. (1)
Итого 26 семантических единиц

Рисунок 5. Исходный текст осмотра с семантической разметкой

6. Обсуждение результатов исследования

Представленные в примере результаты говорят сами за себя. Система транскрибации не справилась с терминами: паренхима (паренем мы, паренхемы), гиперэхогенные (гиперхо генные), не дилятирована (неделя тирована), патологические (потологические).

И при обработке других документов наблюдалось подобное, например «цито» могло транскрибироваться в «пациент», а «хирургический» в «юридический».

Следует отметить, что ошибки транскрибации это известная проблема. ASR система использует две модели акустическую и языковую. Модели взаимодействуют при транскрибации аудио в текст следующим образом: акустическая модель определяет звуки в аудиозаписи, а языковая готовит текстовый материал на основе полученных данных.

Существует возможность до обучать базовую языковую модель (применение техник fine-tuning на медицинских данных), которую использует ASR и работать с определенным конкретными терминами, но

Осмотр почек правая почка расположена обычно размерами 9, запятая 9 на 3, запятая 8 на 3, запятая 8 сантиметров, контуры ровные, чёткие, толщина паренем мы 1 запятая 9 сантиметров, структура паренхемы, почки несколько неоднородная.

Эхогенность обычная в почечном синусе мелкие линейные гиперэхогенные структуры без чётких акустических теней, уплотнённые стенки сосудов, чашечно лоханочная система недели тирована, левая почка несколько опущена размерами 10 и 9.

На 5 запятая 0, на 4 запятая 5 сантиметров, контуры ровные чёткие, толщина паренхемы 1 запятая 8 сантиметров, структура паренхемы несколько неоднородная, эхогенность обычная в сосудистых структурах, в паренхиме.

Киста размером 1, запятая 7 на 1, запятая 3 на 1, запятая 4 сантиметра в почечном синусе мелкие линейные гиперхо генные структуры без чётких акустических теней, уплотнённые стенки сосудов, чашечно лоханочная система, неделя тирована в проекции надпочечников, потологические образования, они определяются.

При узи ангиографии ход основных сосудистых структур почек обычный в обеих почках, во всех сегментах кровоток прослеживается до периферических отделов коркового слоя заключение киста левой почки, рекомендации, контроль, узи патологии.

Отсутствует диагноз киста, почки впервые выявлено

Рисунок 6. Текст осмотра после транскрибации SaluteSpeech

ОСМОТР ПОЧЕК:

Правая почка (состояние и размеры): расположена обычно, размеры 9,9x3,8x3,8 см

Контуры: ровные, чёткие

Толщина паренхимы: 1,9 см

Структура паренхимы почки: несколько неоднородная

В почечном синусе: мелкие линейные гиперэхогенные структуры без чётких акустических теней, уплотнённые стенки сосудов

Чашечно-лоханочная система: не дилятирована

Левая почка (состояние и размеры): несколько опущена, размеры 10,9x5,0x4,5 см

Контуры: ровные, чёткие

Толщина паренхимы: 1,8 см

Структура паренхимы почки: несколько неоднородная

В почечном синусе: мелкие линейные гиперэхогенные структуры без чётких акустических теней, уплотнённые стенки сосудов

Чашечно-лоханочная система: не дилятирована

Проекция надпочечников: патологические образования определяются

УЗ-ангиография: ход основных сосудистых структур почек обычный в обеих почках, во всех сегментах кровоток прослеживается до периферических отделов коркового слоя

Заключение: киста левой почки размером 1,7x1,3x1,4 см

Рекомендации: контроль УЗИ

Патологии: киста левой почки

Рисунок 7. Текст осмотра после обработки YaGPT

это дорогостоящий и продолжительный процесс. Так же существуют

технические и организационные методы улучшения качества, но практика показывает, что в условиях реального лечебного учреждения получить идеальное качество звукового файла невозможно и надо быть готовым к искажениям и ошибкам.

Приводим в таблице 3 ошибки транскрибации наиболее часто встретившиеся в вычислительных экспериментах.

Таблица 3. Статистика по ошибкам транскрибации

Исходный термин	Процент ошибок	Примеры ошибок транскрибации
гастроэзофагеальная	100%	гастроэзофагиальная, Гастроэзофагеназия, Гастроэзофагинальная
ГЭРБ	100%	грп, Герб, ГАРП
дилятирована	100%	недели тирована, Деля тирована, дилатировано, дилятина,
казеозных	100%	кови зные, ковиозных, козеозных, козелозных,
тонзиллитиаз	100%	Танзили тиаз, тони плитас, литиаз, тонзиллит аз,
трабекулярных	100%	Рапи, улярных, траулерных, трабу биокулярных, тробикулярных
тугоэластическая	80%	туго-эластическая, туга эластическая
ЦДК	80%	CDK
эутиреоз	80%	и утериоз, Эу. Тиреоз, Эутиреолоз
инъецирована	60%	инъецированная, инфицирована, инициирован
околоносовых	60%	около носовых
внутрипросветные	40%	внутрепросветные, внутри просветные,

Нейросеть YaGPT исправила все приведенные в примере ошибки транскрибации, а также привела текст к требуемой структуре. И при обработке других документов почти со всеми ошибками транскрибации LLM удавалось справиться.

Более сложной задачей для LLM стала задача приведения текста к требуемой структуре. В приведенном примере описание патологии «В сосудистых структурах в паренхиме киста (1) размерами 1,7×1,3×1,4 см (1)» находилось в описании паренхимы, а в конечном результате эти две семантические единицы были перемещены в заключение. С одной стороны видим, что семантика, связанная с патологией, не была потеряна, но с другой стороны изменилось положение семантических единиц

в структуре документа.

Указанные семантические единицы следовало бы оставить в описании паренхимы, а затем на основании описания сделать вывод о патологии. В полученном результате вывод видится ничем не обоснованным. Хотя семантические единицы не были потеряны, их исключение из описания паренхимы следует считать семантической ошибкой.

Иногда после обработки LLM некоторые семантические единицы полностью терялись. Возможно нейросеть решала, что эти СЕ не вписываются в контекст и структуру документа. Таких случаев в проведенных экспериментах было немного, но они имели место.

Вот пример. В исходном осмотре врача ЛОРа была фраза – «В дополнительном обследовании не нуждается». После обработки LLM поле «План обследования» в структуре документа было оставлено пустым. С одной стороны, врач действительно не назначил дополнительных обследований и можно было бы поле не заполнять, но с другой стороны, высказывание врача, по нашему мнению, обладает большей модальной силой, чем «пустое высказывание» (пустое поле). Данный случай мы расценили как ошибку и потерю LLM семантики.

Мы оставляем за рамками статьи техническую часть реализации решений, проблему программного формирования запросов (prompts) для LLM. Эксперименты показали, что даже «тяжелые» LLM с многими миллиардами параметров ошибаются, и необходимо экспериментально подбирать запросы, контекст и формат шаблонов медицинских документов, чтобы улучшить качество результата от LLM.

Заключение

По итогам исследования можно сделать однозначный вывод, что ожидаемый эффект от применения LLM совместно с системой транскрибации наблюдается. LLM в основном справляется со всеми поставленными ей задачами: исправление ошибок транскрибации, синтез информации из файла транскрибации и шаблона документа, приведение текста к заданной структуре.

Полученные оценки качества двух решений позволяют утверждать, что можно достигать приемлемого качества для применения решения в медицинской практике для автоматизации голосового ввода медицинских данных. Естественно, что окончательное решающее слово в вопросе практического применения останется за врачами.

Исследования должны быть продолжены в части применения шаблонов заполнения документов.

Полученные результаты могут быть полезны и востребованы разработчиками современных интеллектуальных медицинских систем.

Список использованных источников

[1] Малых В. Л., Гулиев Я. И., Калинин А. Н., Колупаев А. В., Юрченко С. Г. *Возможности применения речевого интерфейса и систем автоматической обработки текстов в МИС* // Врач и информационные технологии.– 2014.– № 5.– С. 37–47. ↑198

[2] Sumrak J. *Medical voice recognition: How AI solves terminology problems.*– 2025. ↑199, 200, 201

[3] Kumar K. *Benchmarking Automatic Speech Recognition coupled LLM modules for medical diagnostics.*– 2025.– 7 pp. arXiv 2502.13982v1~[eess.AS] ↑199, 201

[4] Wang D., Zhang S. *Large language models in medical and healthcare fields: applications, advances, and challenges* // Artif. Intell. Rev..– 2024.– Vol. **57**.– id. 299.– 48 pp. ↑201

[5] Гусев А. *Большие языковые модели (LLM) в здравоохранении*, 28 октября.– WEBIOMED.– 2024. ↑201

[6] В. Л. Малых *Системы поддержки принятия решений в медицине* // Программные системы: теория и приложения.– 2019.– Т. **10**.– № 2(41).– С. 155–184. ↑201

[7] *White House Unveils America's AI action plan*, July 23.– The White House.– 2025. ↑201, 203

[8] Adedeji A., Sanni M., Ayodele E., Joshi S., Olatunji T. *The multicultural medical assistant: can LLMs improve medical ASR errors across borders?*– 2025.– 15 pp. ↑201

[9] Малых В. Л., Калинин А. Н., Рудецкий С. В. *Архитектура взаимодействия в медицинской экосистеме* // Программные системы: теория и приложения.– 2024.– Т. **15**.– № 2.– С. 475–492. ↑203

[10] *ГОСТ Р 59921.5-2022. Системы искусственного интеллекта в клинической медицине. Часть 5. Требования к структуре и порядку применения набора данных для обучения и тестирования алгоритмов.*– М.: Российский институт стандартизации.– 2022.– 24 с. ↑205, 206

[11] Дементьев А. В. *Метрики семантических данных* // Молодой ученый.– 2022.– № 24 (419).– С. 48–51. ↑209

Поступила в редакцию 22.10.2025;
одобрена после рецензирования 30.10.2025;
принята к публикации 17.11.2025;
опубликована онлайн 15.12.2025.

Рекомендовал к публикации *д.м.н. Т.В. Зарубина*

Информация об авторах:

Юрий Геннадьевич Сидоров

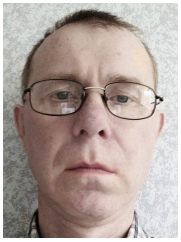
Руководитель проектов ООО «Интерин технологии». Научный интерес в изучении трендов и проведении исследований с открытыми модели (open-source LLM) и автоматизации рабочих процессов с помощью малых, специализированных моделей (small language models) и агентных систем.

0009-0002-6352-5173
e-mail: *sidorov@interin.ru*

Владимир Леонидович Малых

Зав. лабораторией Института программных систем им. А. К. Айламазяна. Научные интересы: Медицинские информационные системы, платформы для МИС, интеграционные профили и личные кабинеты участников медицинских экосистем.

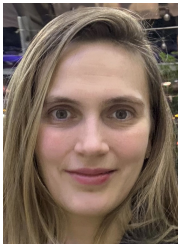
0000-0002-0072-0724
e-mail: *mvl@interin.ru*



Алексей Николаевич Калинин

Ведущий инженер-программист ООО «Интерин технологии», ведущий разработчик семейства МИС Интерин. Научные интересы: Медицинские информационные системы, компоненты разработки МИС на платформе Интерин Альфа, интеграционные профили участников медицинских экосистем.

0000-0002-3607-7400
e-mail: *ank@interin.ru*



Ольга Сергеевна Елистратова

Младший научный сотрудник Института программных систем им. А. К. Айламазяна. Научные интересы: Медицинские информационные системы

0000-0002-8975-3839
e-mail: *ola@interin.ru*

Авторы внесли равный вклад в подготовку публикации.

Декларация об отсутствии личной заинтересованности: *благополучие авторов не зависит от результатов исследования.*

EN RU PROGRAM SYSTEMS: THEORY AND APPLICATIONS ISSN 2079-3316

Research Article INFORMATION TECHNOLOGY IN MEDICINE

UDC 61:007.51:316

10.25209/2079-3316-2025-16-6-197-219

Automatic Speech Recognition coupled LLM

Yurij Gennadievich **Sidorov**[1], Vladimir Leonidovich **Malykh**[2], Aleksey Nikolayevich **Kalinin**[3], Olga Sergeevna **Yelistratova**[4]

[1, 3] Interin Group of Companies, Moscow, Russia
[2, 4] Ailamazyan Program Systems Institute of RAS, Ves'kovo, Russia

Abstract. One of the barriers preventing the widespread use of speech medical data entry in HIS is the insufficient consumer quality of texts obtained after transcription. Not all medical terms and words of the general lexicon are recognized correctly, the coordination of words by gender, number and case is disrupted, the text is not well formatted from the point of view of grammar. All this requires further revision of the text. Another difficult problem is the need to bring the text to the structure of the medical document in HIS. The document structure can be quite complex, contain many elements, and have requirements for the type and format of the data in the structure elements. Speech input can only be partially used to generate a document, and the missing data can be taken from a custom template.

To solve these problems, we propose to use LLM as a corrector of speech transcription results, an integrator of text data and text data from a template, and a structurer of the resulting data. The paper proposes a solution architecture for the input of speech medical data based on the composition of the transcription system and LLM. A methodology for conducting solution tests is proposed, including the preparation of a dataset and a metric for calculating the quality of the solution. The implementation of the solution based on a free and proprietary component is described.

The results can be used in the development and evaluation of AI systems used for speech data input, and not only in medicine. (*In Russian*).

Key words and phrases: medical informatics, medical information systems, speech recognition, medical voice recognition, LLM

2020 *Mathematics Subject Classification*: 94A05; 92C50, 93Bxx

For citation: Yurij G. Sidorov, Vladimir L. Malykh, Aleksey N. Kalinin, Olga S. Yelistratova. *Automatic Speech Recognition coupled LLM*. Program Systems: Theory and Applications, 2025, **16**:6(71), pp. 197–219. (*In Russ.*). https://psta.psiras.ru/read/psta2025_6_197-219.pdf

References

[1] V. L. Malyx, Ya. I. Guliev, A. N. Kalinin, A. V. Kolupaev, S. G. Yurchenko. "The possibility of using a speech interface and automatic text processing systems in MIS", *Vrach i informacionnye texnologii*, 2014, no. 5, pp. 37–47 (in Russian).

[2] J. Sumrak. *Medical voice recognition: How AI solves terminology problems*, 2025.

[3] K. Kumar. *Benchmarking Automatic Speech Recognition coupled LLM modules for medical diagnostics*, 2025, 7 pp. arXiv 2502.13982v1~[eess.AS]

[4] D. Wang, S. Zhang. "Large language models in medical and healthcare fields: applications, advances, and challenges", *Artif. Intell. Rev.*, **57** (2024), id. 299, 48 pp.

[5] A. Gusev. *Large language models in healthcare fields*, 28 oktyabrya, WEBIOMED, 2024 (in Russian).

[6] Malyx V. L.. "Decision support systems in medicine", *Program Systems: Theory and Applications*, **10**:2(41) (2019), pp. 155–184 (in Russian).

[7] *White House Unveils America's AI action plan*, July 23, The White House, 2025.

[8] A. Adedeji, M. Sanni, E. Ayodele, S. Joshi, T. Olatunji. *The multicultural medical assistant: can LLMs improve medical ASR errors across borders?* 2025, 15 pp.

[9] V. L. Malyx, A. N. Kalinin, S. V. Rudeckij. "Architecture of interaction in the digital medical ecosystem", *Program Systems: Theory and Applications*, **15**:2 (2024), pp. 475–492 (in Russian).

[10] *GOST R 59921.5-2022. Artificial intelligence systems in clinical medicine. Part 5. Requirements for the structure and order of application of a dataset for training and testing algorithms*, Rossijskij institut standartizacii, M., 2022, 24 pp. (in Russian).

[11] A. V. Dement'ev. "Metrics of semantic data", *Molodoj uchenyj*, 2022, no. 24 (419), pp. 48–51 (in Russian).