

Модуль речевого ввода данных с LLM-корректором

Руководство пользователя

Работа на здоровье

INTERIN
ТЕХНОЛОГИИ

2026 г.

Модуль речевого ввода данных с LLM-корректором Руководство пользователя

Документ разработан ООО «Интерин технологии» (©).

Все права защищены. Никакая часть настоящего документа не может быть воспроизведена или передана в какой бы то ни было форме и какими бы то ни было средствами, будь то электронные или механические, включая фотокопирование, запись на магнитный носитель, электронную почту и публикации в Интернет, если на то нет письменного разрешения автора.

Контактная информация

ООО «Интерин технологии»
Web: www.interin.ru
E-mail: info@interin.ru
Тел./факс: +7 (495) 220 82 35

1 ОБЩИЕ СВЕДЕНИЯ

1.1 О СИСТЕМЕ

Полное наименование: Модуль речевого ввода данных с LLM-корректором.

Краткое наименование: Модуль.

Разработчик и правообладатель: ООО «Интерин технологии»

Модуль представляет собой программу для транскрибации речевого потока в текст с последующей коррекцией ошибок транскрибации с помощью больших языковых моделей (LLM). На входе программа получает речевой поток, на выходе выдает транскрибированный и исправленный текст в формате медицинского документа..

Эксплуатационным назначением модуля является применение в здравоохранении для формирования специализированным программным обеспечением медицинских документов заданной структуры при голосовом вводе данных.

Одним из барьеров, препятствующих широкому распространению голосового ввода медицинских данных в медицинских информационных системах (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 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.

Но речевой ввод может лишь частично использоваться для формирования документа, а недостающие данные могут быть взяты из пользовательского шаблона. Для решения указанных проблем следует использовать LLM в качестве корректора результатов транскрибации речи, интегратора речевых и текстовых данных из шаблона и формирователя структуры результирующих данных.

1) **Коррекция результатов транскрибации.**

Из-за ошибок распознавания медицинских терминов и слов общего лексикона, нарушений грамматического согласования и слабого форматирования текст после транскрибации требует доработки. ПО должно автоматически исправлять эти недостатки с использованием большой языковой модели (LLM). Конкретные задачи коррекции:

- точное распознавание и написание медицинских терминов;
- грамматическая проверка и исправление ошибок согласования (род, число, падеж);
- приведение текста к единому стилю и улучшение его форматирования.

2) **Структурирование данных.**

ПО должно преобразовывать свободный текст, полученный после коррекции, в строго заданную структуру медицинского документа МИС. Структура может включать множество элементов с жёсткими требованиями к типу данных (текст, число, дата, код) и их формату.

3) Интеграция данных из разных источников.

Поскольку в ходе речевого ввода информация может быть представлена не полностью, ПО должно дополнять транскрибированные данные недостающей информацией из заранее подготовленного пользовательского шаблона. LLM должна выступать в роли интегратора, объединяя речевые и шаблонные данные в единый структурированный документ.

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. Обратное отображение этой структуры в дерево тривиально при условии, что все узлы структуры имеют уникальные имена. Именно такие шаблоны структуры документа должны использоваться при реализации.

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: 1) оставлять неизменным (insert); 2) удалять данные заполнения (delete); объединять данные заполнения с дополнительными данными вводимыми пользователями (append); могут отредактировать данные заполнения (edit).

Модуль речевого ввода должен обрабатывать полученный текст и шаблон, осуществляя «умный» синтез данных из речевого ввода с данными из шаблона заполнения.

1.2 НЕОБХОДИМАЯ СРЕДА ФУНКЦИОНИРОВАНИЯ

Для функционирования серверной части Модуля речевого ввода необходимо следующее системное программное обеспечение:

- Системы на базе Debian: Debian 12 и выше Ubuntu 24 и выше.
 - Для функционирования клиентской части Модуля речевого ввода на рабочей станции пользователя должен использоваться следующий набор системного программного обеспечения:
- 1) Веб-браузеры актуальных версий (любой из списка):
 - Яндекс.Браузер;
 - Спутник;
 - Microsoft Edge;
 - Google Chrome;
 - Mozilla Firefox;
 - Opera.

Используемое стороннее ПО:

Библиотека	Версия ¹	Лицензия	Примечания
Aiofiles	25.1.0	MIT	Стандартная разрешительная лицензия

¹ Или последние на момент выпуска продукта стабильные версии

Библиотека	Версия ¹	Лицензия	Примечания
Aiohttp	3.13.3	Apache 2.0	Разрешительная, с явным указанием патентных прав
aiohttp-cors	0.8.1	MIT	Совместима с лицензией aiohttp
Av (PyAV)	16.1.0	BSD 2-Clause	Разрешительная
boto3	1.42.57	Apache 2.0	Стандартная для AWS SDK
ctranslate2	4.7.1	Apache 2.0	Совместима с другими компонентами ML
faster-whisper	1.2.1	MIT	Основан на Whisper от OpenAI (MIT)
Huggingface_hub	1.4.1	Apache 2.0	
Httpx	0.28.1	BSD 3-Clause	Разрешительная
langchain-classic	1.0.1	MIT	Разрешительная
langchain-community	0.4.1	MIT	Аналогично langchain-classic
langchain-core	1.2.16	MIT	
langchain-text-splitters	1.1.1	MIT	
Langsmith (SDK-библиотека для Python)	0.7.7	MIT	Свободна для использования
Numba	0.64.0	BSD 2-Clause	
Onnxruntime	1.24.2	MIT	Разрешительная
openai-whisper	20231117	MIT	
Orjson	3.11.7	Apache 2.0	Разрешительная лицензия
Pydantic	2.12.5	MIT	
pydantic-settings	2.13.1	MIT	
Pydub	0.25.1	MIT	Разрешительная
python-dotenv	1.2.1	BSD 3-Clause	
PyYAML	6.0.3	MIT	
Requests	2.32.5	Apache 2.0	
Rich	14.3.3	MIT	
Scipy	1.17.0	BSD 3-Clause	
Sounddevice	0.5.5	MIT	
SQLAlchemy	2.0.47	MIT	
Srt	3.5.3	MIT	Разрешительная
Tenacity	9.1.4	Apache 2.0	
Tiktoken	0.12.0	MIT	
Tokenizers	0.22.2	Apache 2.0	
Torch (PyTorch)	2.10.0	BSD 3-Clause	Разрешительная
Tqdm	4.67.3	MIT	
Typer	0.24.1	MIT	
Vosk	0.3.45	Apache 2.0	Разрешительная

Библиотека	Версия ¹	Лицензия	Примечания
Websockets	16.0	BSD 3-Clause	
Модель Qwen	2.5-32B-Instruct (квантованная 4-8 бит)	Apache 2.0	Разрешительная
Модели Vosk	например, vosk-model-ru-0.22, vosk-model-small-ru-0.22 и другие	Apache 2.0	Разрешительная
Модели Whisper	например, whisper-small, whisper-large-v3 и другие	MIT	Разрешительная

Модуль речевого ввода должен функционировать на технических средствах с характеристиками не ниже следующих.

Сервер

Характеристика	Значение
CPU	От Intel Xeon 2.0 ГГц 8 ядер
GPU	От RTX 4090 24 ГБ
RAM	От 32 ГБ (рекомендуется 64 ГБ для комфортной работы с Qwen2.5-32B)
Дисковое пространство	От 500 ГБ (RAID 10)

Рабочие станции

Параметры и характеристики классической рабочей станции:

Устройство	Характеристики
CPU	От 2 ГГц
RAM	От 4 ГБ
Дисковое пространство	От 80 ГБ
Видеокарта	Поддержка видеорежима 1280 x 1024 и выше
Монитор	От 19.0" LCD монитор

Принтер (опционально)

Сетевая инфраструктура

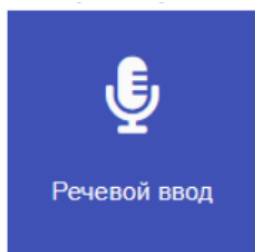
Клиенты взаимодействуют с серверами на скорости не ниже 100 Мбит/сек.

2 ИСПОЛЬЗОВАНИЕ МОДУЛЯ

Модуль работает в составе используемой Медицинской организацией МИС.

Далее будет продемонстрировано использование Модуля на примере ультразвукового исследования мочевого пузыря. Использование Модуля для формирования медицинских документов в ходе других исследований или врачебных приемов аналогично.

- 1) Для запуска демонстрации возможностей модуля речевого ввода следует кликнуть по плитке:



- 2) В открывшейся форме демонстрации возможностей Модуля следует кликнуть по кнопке «Речевой ввод». При работе в составе МИС такая кнопка будет появляться для формирования тех медицинских документов, для которых возможен данный способ формирования документов (загружены соответствующие шаблоны, разработаны соответствующие промпты для LLM).

Речевой ввод

Ультразвуковое исследование мочевого пузыря

Протокол исследования
1

Речевой ввод

Контур	<input type="text"/>
Форма	<input type="text"/>
Эхогенность	<input type="text"/>
Содержимое	<input type="text"/>
Объём	<input type="text" value="куб.см"/>
Толщина стенки	<input type="text" value="мм"/>
Эхоструктура	<input type="text"/>
Объёмные образования	<input type="text"/>
Объём остаточной мочи	<input type="text" value="куб.см"/>
Патологические изменения	<input type="text"/>
Дополнительная информация	<input type="text"/>
Заключение*	<input type="text"/>
Рекомендации	<input type="text"/>

> Отладочная информация

- 3) При запуске речевого ввода откроется форма ввода данных. Для начала ввода следует включить микрофон, кликнув по кнопке «1. Включить микрофон» – форма перейдет в состояние «Запущено» (микрофон включен). В ходе комментариев пользователя (врача, в данном случае предполагается описание врачом «вслух» увиденного на экране УЗИ-

аппарата при исследовании мочевого пузыря) транскрибированный текст выводится в окне «Результаты транскрибации».

В демонстрации используется протокол УЗИ Мочевого пузыря, в ходе которой рекомендуется произнести значения следующих параметров (обычно врачи описывают их в протоколе данного исследования):

Контуры: ровные / волнистые / не ровные / другое ___

Форма: обычная / симметричная / овальная / округлая / не правильная / не симметричная / вытянутая / другое ___

Эхогенность: не изменена / повышена / понижена / смешанная/ другое ___

Содержимое: анэхогенное / другое ___

Объем: ___ мл

Толщина стенки: ___ мм

Эхоструктура: однородна/ диффузно неоднородна/ неоднородна за счёт ___

Объемные образования: не лоцируются / конкремент d ___ мм / взвесь / другое ___

Патологические изменения: не обнаружены / другое

Порядок произнесения полей не имеет значения, допускается ведение сторонних разговоров, например инструкции пациенту.

Пример текста: Секунду, сейчас будет чуть-чуть холодно, форма обычная, контуры ровные, наполнен 250 мл, толщина стенки 3 мм, не дышите, так нормально, в просвете подвижная структура гиперэхогенная 35 на 20, дающая акустическую тень. Вот салфетка вытирайтесь.

- 4) Когда пользователь считает ввод данных завершённым, ему следует кликнуть кнопку «2. Выключить микрофон». Учитывая задержки передачи данных по сети, рекомендовано выключать микрофон через 3 - 5 сек. после окончания речевого ввода. Форма перейдет в состояние «Остановлено».
- 5) Далее следует выставить желательную LLM в окне «Языковая модель» и отправить результат для обработки, кликнув кнопку «3. Отправить в LLM на коррекцию».

- 6) После преобразований результат обработки транскрибированного текста будет выведен в окне «Результат обработки» в виде json.

Речевой ввод данных с LLM-корректором

1. Включить микрофон 2. Выключить микрофон

Состояние: **Остановлено**

Результат транскрибации речи

секундочку сейчас будет чуть-чуть холодно
форма обычная
контуры ровные
наполнен двести пятьдесят миллилитров
толщина стенки три миллионер миллиметра
не дышите
так нормально
в просвете подвижная структура гиперэхогенная двадцать пять на тридцать дающая акустическую тень
бот салфетка вытирайтесь

Языковая модель local 3. Отправить в LLM на коррекцию

7 Результат обработки

```

{
  "PROTOCOL": {
    "FORM": "обычная",
    "VOLUME": "двести пятьдесят миллилитров",
    "CONTENT": "",
    "ADD_DATA": "",
    "CONTOURS": "ровные",
  }
}

```

ОЧИСТИТЬ

- 7) Справа от формы ввода показана инструкция по выполнению действий. После заполнения поля «Результат обработки» следует нажать кнопку «4. Вставить в протокол», расположенную под инструкцией.

Речевой ввод данных с LLM-корректором

1. Включить микрофон 2. Выключить микрофон

Состояние: **Остановлено**

Результат транскрибации речи

секундочку сейчас будет чуть-чуть холодно
форма обычная
контуры ровные
наполнен двести пятьдесят миллилитров
толщина стенки три миллионер миллиметра
не дышите
так нормально
в просвете подвижная структура гиперэхогенная двадцать пять на тридцать дающая акустическую тень
бот салфетка вытирайтесь

Языковая модель local 3. Отправить в LLM на коррекцию

Результат обработки

```

{
  "PROTOCOL": {
    "FORM": "обычная",
    "VOLUME": "двести пятьдесят миллилитров",
    "CONTENT": "",
    "ADD_DATA": "",
    "CONTOURS": "ровные",
  }
}

```

Инструкция

- Включить микрофон
Продиктовать текст протокола.
В демонстрации используется протокол УЗИ Мочевого пузыря, рекомендуется произнести следующие значения полей:
Контуры: ровные / волнистые / не ровные / другое ____
Форма: обычная / симметричная / овальная / округлая / не правильная / не симметричная / выгнутая / другое ____
Эхогенность: не изменена / повышена / понижена / смешанная / другое ____
Содержимое: анэхогенное / другое ____
Объем: ____ мл
Толщина стенки: ____ мм
Эхоструктура: однородная / диффузно неоднородная / неоднородная за счёт ____
Объемные образования: не локализуются / конкретен в ____ мм / вазель / другое ____
Патологические изменения: не обнаружены / другое ____
Порядок произнесения полей не имеет значения, допускается ведение стартовых разговоров, например инструкции пациенту.
Пример текста: Секунду, сейчас будет чуть-чуть холодно, форма обычная, контуры ровные, наполнен 250 мл, толщина стенки 3 мм, не дышите, так нормально, в просвете подвижная структура гиперэхогенная 35 на 20, дающая акустическую тень. Бот салфетка вытирайтесь.
- Выключить микрофон (учитывая задержки передачи данных по сети, рекомендовано выключать микрофон через 3 - 5 сек. после окончания речевого ввода).
- Выбрать языковую модель и нажать кнопку "3. Отправить в LLM для обработки"
- После заполнения поля "Результат обработки" нажать кнопку "4. Вставить в протокол"

8 4. Вставить в протокол Закрыть

- 8) Во время обработки текста и формирования документа (протокола исследования) на экран будет выдано сообщение о ходе процесса («Операция выполняется. Пожалуйста, подождите»).
- 9) В результате произведенных действий будет сформирован протокол исследования.

Ультразвуковое исследование мочевого пузыря

Протокол исследования		Речевой ввод
Контур	ровные	
Форма	симметричная	
Эхогенность	не изменена	
Содержимое	анэхогенное	
Объем	250 миллилитров	
Толщина стенки	4 миллиметра	
Эхоструктура	однородна	
Объемные образования	не лоцируется	
Объем остаточной мочи	куб. см	
Патологические изменения	не обнаружено	
Дополнительная информация		
Заключение*		
Рекомендации		