



TailSense

AI помощник для диагностики заболеваний с/х животных
ИТ стипендия РСХБ цифра

Гаттаров Тимур и Маталыцкий Иван
МФТИ

30.10.2025

Содержание

1. Постановка задачи
2. Актуальность проблемы
3. Архитектура решения
4. Архитектура RAG системы
5. Бэнчмарки
6. Архитектура AI агента
7. Backend и архитектура
8. UI/UX и сценарии использования
9. Вызовы и ограничения
10. Заключение



Репозиторий с проектом
на github

Постановка задачи

Цель проекта: Разработать AI справочник болезней сельхозживотных с/х:

- Диагностикой по симптомам
- Рекомендациями по лечению
- Мерами профилактики
- Поиском по виду животного

Ключевые требования:

- Удобный и понятный интерфейс
- Актуальная и достоверная информация
- Быстрый поиск



Актуальность проблемы

Фермеры сталкиваются с множеством проблем при заболевании животных.

При возникновении болезни у скота диагностика затруднена так как:

- Доступность ветеринарных специалистов в регионах ограничена
- Отсутствует информация и инструменты для предварительной оценки состояния животных

Это приводит к:

- Задержкам в оказании помощи
- Рискам распространения заболеваний в поголовье
- Экономическим рискам для фермеров в связи с падежом скота

Традиционный подход и его ограничения

Ожидаемое решение: классическая база данных с поиском

Стандартный подход:

- Реляционная база данных
- Поиск по ключевым словам
- Жёсткая структура данных
- Статический контент
- Ручное обновление информации

В итоге фермер получает либо слишком общий ответ, либо система вообще не может понять суть проблемы

Основные проблемы:

- **Не понимает контекст:**
"температура + кашель" "кашель + температура"
- **Требует точных формулировок:**
"лихорадка" "повышенная температура"
- **Ограниченная функциональность:**
Только поиск, нет диалога
- **Сложность поддержки:** Ручное добавление новых данных

Наше решение

TailSense — прототип умного AI помощника, который всегда под рукой

Что это такое:

- Цифровой ветеринарный справочник с искусственным интеллектом
- Мгновенная диагностика по симптомам в чате
- Работает 24/7 на телефоне и компьютере

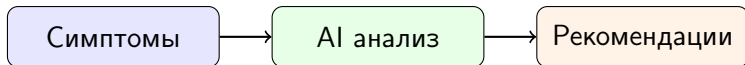
Как это может помочь:

- Диагностика за несколько минут вместо часов ожидания
- Профилактика эпидемий
- Продлевает жизни животных и деньги фермеров

Просто опишите симптомы — получите понятные рекомендации

Обзор работы проекта

Мы создали AI-помощника, который объединяет ветеринарные знания с передовыми технологиями ИИ.



В основе системы — технология **RAG (Retrieval-Augmented Generation)**. Мы преобразуем справочники в векторные эмбединги через **OpenAI text-embedding-3-small** и храним их в векторной БД **Qdrant**.

При описании симптомов система выполняет семантический поиск, находит релевантную информацию и передаёт её LLM для генерации рекомендаций. Если система не находит релевантную информацию, то у LLM есть доступ в интернет.

Используемая литература и источники

В нашем прототипе агентная система имеет доступ к информации из следующих книг:



Кодекс здоровья наземных животных МЭБ (2021)

Международное эпизоотическое бюро



Актуальные инфекционные болезни крупного рогатого скота

Современные методические рекомендации (2021)



Внутренние болезни животных

Шербаков Г.Г., Ковалев С.П. (6-е издание)



Справочник ветеринарного специалиста

Ханников А.А.

Охват знаний: Для полноценного AI помощника этого, безусловно, мало, но текущего набора достаточно для демонстрации ключевых возможностей системы.

Архитектура системы

Data Layer

- Векторная БД
- Web search

AI Layer

- RAG Pipeline
- LLM Orchestration
- Agents

Application Layer

- API Gateway
- Web UI

Технологический стек RAG системы



OpenAI text-embedding-3-small

- Качественные эмбединги
- Оптимизированная стоимость
- Высокая скорость обработки



Qdrant Cloud

- Облачное векторное хранилище
- Быстрый семантический поиск
- Масштабируемость

Загрузка данных в RAG реализована в этом ноутбуке

Процесс загрузки данных в RAG систему

Data Ingestion Pipeline: Подготовка знаний для AI

Основные этапы:

1. Загрузка PDF

- Ветеринарные справочники
- Русский/английский языки
- Сохранение структуры

2. Семантическое разбиение

- Чанк: 1000 токенов
- Перекрытие: 10%

3. Создание эмбеддингов

- OpenAI embedding-3-small
- 1536 измерений
- Для семантического поиска

Метаданные:

- Название книги
- Автор и год издания
- Номер страницы

Процесс извлечения информации из RAG

Query Processing Pipeline: Интеллектуальный поиск ответов

Обработка запроса:

1. Анализ интенга

- Тип запроса
- Ключевые сущности
- Вид животного

2. Семантический поиск

- Векторный поиск
- Фильтрация по метаданным

3. Переранжирование

- По релевантности
- Контекстуальная важность
- Фильтрация дубликатов

Подготовка ответа:

1. Контекстуализация

- Объединение чанков
- Связный контекст
- История диалога

2. Генерация

- LLM + RAG контекст
- Ссылки на источники
- Рекомендации

Пример работы RAG системы

Запрос пользователя

"Что такое лептоспироз?"

Ответ системы

Диагноз: Лептоспироз — заболевание, вызываемое лептоспирами

Симптомы: Лихорадка, поражение почек, печени, нервной системы

Лечение: Антибиотики, специфическая сыворотка, симптоматическая терапия

Источники:

- SHerbakov_G.G. et al. "Внутренние болезни животных"(стр. 241)
- SHerbakov_G.G. et al. "Внутренние болезни животных"(стр. 434)

Ключевое преимущество:

Как видно, каждое утверждение имеет ссылку на источник, это позволяет бороться с галлюцинациями LLM, так как информацию можно легко и быстро перепроверить в источниках.

Бэнчмарк качества системы

Наша задача относится к классу **close-ending** задач, то есть существует конечное множество корректных диагнозов и методов лечения. Это фундаментальное свойство позволяет нам создать набор тестовых вопросов с заранее известными ответами.

Для объективной оценки качества работы нашего AI-агента мы разработали baseline тестирования. Сначала мы задаем вопрос нашему агенту, который был обучен на специализированной ветеринарной литературе. Затем полученный ответ вместе с исходным вопросом передается на валидацию модели GPT-4o – **judge**.

После этого **judge** сравнивает ответ агента с верным вариантом и присваивает ему одну из трех категорий оценки: правильный ответ, частично правильный или неправильный. Такой подход позволяет получить примерные метрики точности системы.

Результаты тестирования: Стандартные запросы

Тестирование на корректных вопросах из базы знаний

Система демонстрирует хорошие результаты при работе с проверенными диагностическими сценариями, которые полностью соответствуют изученной литературе.

- Правильных ответов: 16 (55.2%)
- Частично правильных: 13 (44.8%)
- Неправильных: 0 (0.0%)
- Общая точность: 77.6%

Эти результаты показывают, что система успешно усвоила основную базу знаний и может довольно надежно работать с типичными случаями из ветеринарной практики. Частично правильные ответы обычно содержат верную диагностику, но неполные рекомендации по лечению.

Результаты тестирования: Правдоподобные заблуждения

Тестирование на вопросах, похожих на правду

При столкновении с вопросами, которые звучат разумно, но содержат неточную информацию или ту которой нет в нашей базе знаний, производительность системы снижается.

- Правильных ответов: 23 (76.7%)
- Неправильных: 7 (23.3%)

Это наиболее сложная категория, где система иногда принимает ложные предпосылки за истинные. Неправильные ответы возникают, когда AI не распознает скрытые заблуждения в формулировке вопроса.

Результаты тестирования: Абсурдные запросы

Тестирование на бредовых вопросах

Система показала отличные результаты при обработке явно абсурдных или ненаучных запросов, уверенно их отклоняя.

- **Правильных ответов:** 15 (100.0%)
- **Неправильных:** 0 (0.0%)

Высокая точность демонстрирует надежность системы в фильтрации непроверенных методов и предотвращении распространения ложной информации.

Анализ результатов и доступ к бэнчмарку

Ключевые выводы по всем категориям

Система демонстрирует разную производительность в зависимости от типа запроса: отличную при работе с абсурдными вопросами и хорошую со стандартными случаями и правдоподобными заблуждениями.

- **Стандартные вопросы:** 77.6% точности
- **Правдоподобные заблуждения:** 76.7% точности
- **Абсурдные запросы:** 100.0% точности

Процент неточностей можно сократить, если мы расширим базу знаний.

GitHub: Ноутбук с полным бэнчмарком

Инструменты для построения агента



LangGraph

- Оркестрация многошаговых процессов
- Управление состоянием агента
- Сложные workflows



LangFlow

- Визуальное проектирование цепочек
- Быстрое прототипирование
- Интуитивный интерфейс



DuckDuckGo

- Веб-поиск актуальных данных
- Быстрые ответы

Архитектура AI агента

Plan → Execute → Reflect → Iterate

Процесс работы:

1. **Plan** — анализ запроса и планирование диагностики
2. **Execute** — использование инструментов для сбора данных
3. **Reflect** — анализ результатов и проверка противоречий
4. **Iterate** — повторение цикла при необходимости

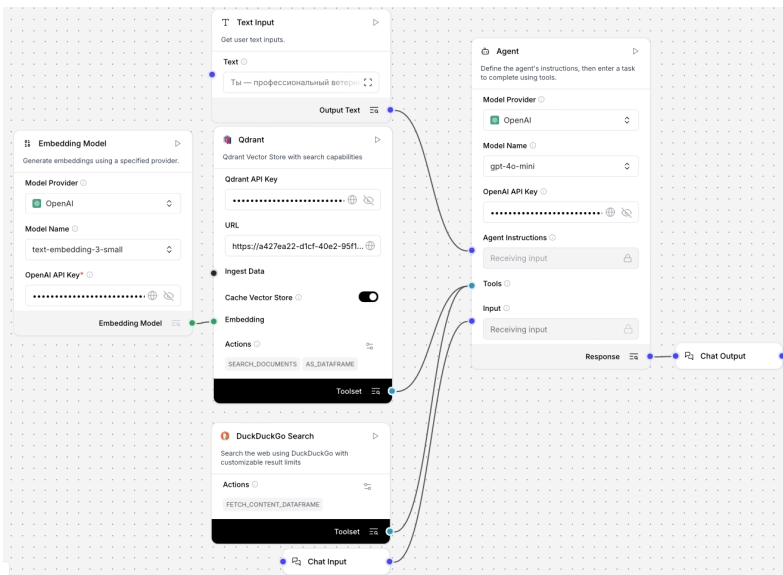
Инструменты агента:

- Векторный поиск в базе знаний
- Веб-поиск актуальных данных

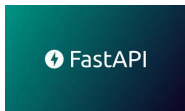
Ключевые преимущества:

- Рекомендации с ссылками на источники
- Адаптация под сложность случая
- Прозрачный процесс принятия решений

Архитектура AI агента

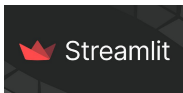


Backend, Frontend, DevOps



Backend

- FastAPI - асинхронный фреймворк, поддерживающий высокую производительность



Frontend

- Мы выбрали Streamlit, так как среди нас нет фронтэнда разработчиков, а благодаря Streamlit можно не сложно создать красивый и мощный интерфейс



DevOps

- Docker для контейнеризации
- Docker Compose для оркестрации

Главный экран приложения




 **TailSense**

Добро пожаловать в TailSense!

TailSense — это ветеринарный помощник на основе ИИ, который поможет вам:

 Быстро оценить симптомы вашего питомца  Получить рекомендации по первой помощи

 **Важно:** Этот сервис не заменяет профессиональную ветеринарную помощь. При серьезных симптомах обязательно обратитесь к ветеринару!

 Начать

>  Дополнительная информация

Элементы:

- Логотип
- Небольшое описание проекта и его возможностей
- Четкий дисклеймер
- Кнопка "Начать"

Выбор животного и симптомов



Настройка консультации

Пожалуйста, выберите вашего питомца и укажите симптомы для получения персональных рекомендаций.



Выберите животное

Выберите из популярных животных:



Собака



Кошка



Птица



Кролик



КРС (корова/бык)



Лошадь

Или введите своё животное:

Название животного

Например: хомяк, игуана, коза...

Выбрано: Кошка

- Выпадающий список видов животных
- Ручной ввод при отсутствии нужного вида
- Обширная база симптомов с поиском
- Кнопка "Назад" для изменения выбора
- Кнопка "К чату" для перехода к диагностике

Выбор животного и симптомов

2 Укажите симптомы

Общие состояние

☒ Повышенная температура

☐ Пониженная температура

☐ Вялость, апатия

☐ Потеря аппетита

☐ Повышенная жажда

Дыхательная система

☐ Кашель

☐ Чихание

☐ Затрудненное дыхание

☒ Хрипы

☐ Выделения из носа

Пищеварение

☐ Рвота

☐ Диарея

☐ Запор

☒ Отказ от еды

☐ Повышенный аппетит

Внешние признаки

☐ Выделения из глаз

☐ Слезотечение

☐ Зуд, расчесы

☐ Выпадение шерсти

☒ Изменение цвета слизистых

Поведение

☐ Беспокойство

☐ Агрессивность

☐ Хромота

☐ Дрожи

☐ Скрытность

Добавить собственный симптом:

Опишите симптом

Например: странные звуки, необычное поведение...

Добавить собственный симптом:

Опишите симптом

Например: странные звуки, необычное поведение...

☒ Выбранные симптомы:

1. Повышенная температура

2. Хрипы

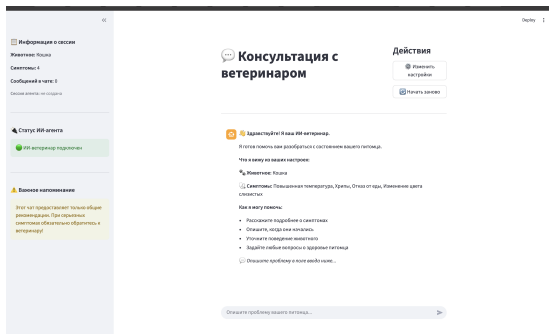
3. Отказ от еды

4. Изменение цвета слизистых

Назад

Перейти к чату

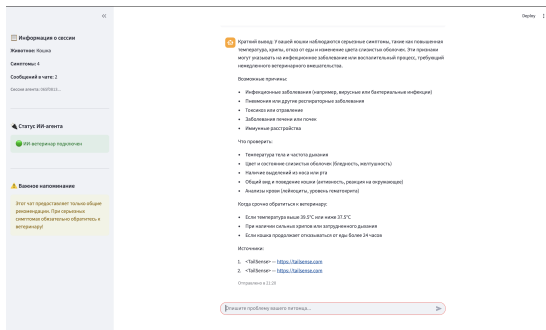
Интерфейс чата с AI-ветеринаром



Элементы интерфейса:

- **Информация о сессии:**
 - Животное и количество симптомов
 - Количество сообщений в чате
- **Четкий дисклеймер**

Структура ответа AI-ветеринара



Компоненты ответа:

- Краткий вывод и предварительный диагноз
- Возможные причины и рекомендации
- Показания срочного обращения к ветеринару
- Источники информации для проверки

Ключевые вызовы и ограничения

Ответственность

- Медицинские риски: ошибки влекут за собой потери животных
- Юридические аспекты: TailSense помощник ветеринару, а не его замена

Меры защиты

- Четкие дисклеймеры
- Многоуровневая проверка (coming soon)
- Мониторинг точности (coming soon)

Будущее развитие

Расширение базы знаний

- Интеграция с государственными ветеринарными базами
- Партнерства с исследовательскими институтами
- Регулярное обновление протоколов лечения
- Добавление региональных особенностей заболеваний

Технологические улучшения

- Мультимодальные модели: анализ фото и видео симптомов
- Automated testing: автоматизированное тестирование точности (наш бэнчмарк является демо версией этого)
- Continuous learning: постоянное обучение на новых данных
- Добавление различных инструментов агенту

Automated testing

В виду высоких рисков, связанных с ветеринарной диагностикой, мы уделяем особое внимание обеспечению точности и надежности нашей системы. Мы понимаем, что каждая ошибка может иметь серьезные последствия для здоровья животных и экономики фермерских хозяйств.

Для достижения высокой точности мы **предлагаем** проводить расширенное тестирование агента на большом объеме данных (с участием LLM as a judge), включая разнообразные заболевания, симптомы и клинические случаи. Перед внедрением обновлений следует выполнять комплексный аудит безопасности, проверяющий как и медицинскую точность, так и устойчивость системы к различным сценариям использования.

После успешного прохождения тестирования и аудита безопасности мы фиксируем версии базы знаний и модели, что обеспечивает стабильность и детерминированность работы системы. Любые последующие обновления будут проходить такой же строгий процесс проверки, гарантируя постоянное поддержание высоких стандартов качества и безопасности.

Перспективы развития: собственная LLM

Для дальнейшего повышения точности и снижения зависимости от внешних сервисов мы рассматриваем возможность разработки собственной языковой модели.

В качестве основы планируется использовать открытые китайские LLM, которые показали хорошие результаты в медицинских и технических областях.

Модель будет дообучена на нашей специализированной базе ветеринарных знаний, что позволит достичь максимальной точности в диагностике заболеваний сельскохозяйственных животных.

Такой подход обеспечит нам полный контроль над моделью, возможность тонкой настройки под конкретные задачи и значительное снижение эксплуатационных затрат в долгосрочной перспективе.

Наша миссия и будущее проекта

- На сегодняшний день в Российском АПК низкий уровень цифровизации, поэтому наша цель — помочь фермерам, особенно из регионов, снизить экономические потери от заболеваний животных и сделать современные технологии доступными для малых хозяйств.
- Мы построили систему с открытой архитектурой, предоставив API для интеграции, чтобы любой разработчик мог подключиться к нашей платформе. Модульная структура позволяет легко расширять функционал, а открытые компоненты обеспечивают прозрачность и возможность совместного развития.

Спасибо за внимание!

Контакты Гатаров Тимур: gattarov.ta@phystech.edu | +7 (903) 225-02-05 | tg: @sugartatar

Контакты Маталыцкий Иван: matalytskii.iv@phystech.edu | +7 (995) 755-96-82 | tg: @zxxxcса