

# **ИИ - ассистенты для улучшения клиентского сервиса**

Эффективные клиники сталкиваются с взаимосвязанными сервисными проблемами:

- очереди на ресепшн в час-пик,
- очереди в колл-центре,
- сложные услуги и множество способов предоставления,
- высокий стресс административного персонала,
- издержки в связи с дефицитом кадров на рынке,
- **необходимость предоставлять сервис, который увеличивает лояльность.**

**Эксперимент:  
повышение лояльности за счет  
повышения заработной платы  
персонала**

**В ходе эксперимента, проведенного в многопрофильной клинике, был применен двухэтапный подход**

1. Увеличение ФОТ на 30% => привлечение более компетентных и мотивированных сотрудников. Потребовалось увеличить зарплату 50 сотрудникам с 40 тыс. ₽ (общий фонд зарплат: 2 млн. ₽) до 60 тыс. ₽, что составило 2,6 млн. ₽, то есть на 600 тыс. ₽ больше.
2. Усиление найма и ротации персонала для предотвращения профессионального выгорания. Это включало своевременную замену, продвижение по карьерной лестнице или увольнение, а также увеличение затрат на HR-поддержку на сумму от 100 до 300 тыс. ₽ в зависимости от размера филиала.

## Результаты эксперимента

Повышение затрат не дало нужного эффекта в связи с:

- дефицитом кадров на рынке труда,
- необходимостью длительно обучать персонал (до 3х месяцев) силами другого квалифицированного персонала,
- дорогостоящие сотрудники тоже выгорают,
- увеличением уровня хаоса в связи с появлением новых сотрудников.
- нежеланием управленцев проводить анализ сырых корпоративных данных, искать корреляции и делать выводы.

Для забора мазка из зева на **Кандиды (Candida albicans)** необходимо придерживаться следующих правил:

**Подготовка пациента:**

За 2-3 часа до забора материала необходимо воздержаться от употребления пищи, питья, не чистить зубы, не полоскать рот, не жевать жевательную резинку и не курить.

**Правила забора:**

- Используется стерильный ворсинчатый зонд.
- Соскоб берется с задней стенки глотки с помощью вращательных движений зонда.
- После взятия материала зонд погружается в эппендорф с транспортной средой.
- Зонд обламывается, в эппендорфе остается ворсинчатая часть зонда.
- Эппендорф закрывается и маркируется в боковом окне (не на крышке).

**Подготовка к исследованию:**

Анализ следует сдавать до начала антибактериальной терапии. Если терапия уже была проведена, сдача анализа должна быть не ранее чем через 3 недели после окончания приема антибиотиков.

Эти меры помогают получить точные результаты при исследовании на наличие Кандиды.

19:50

## Альтернативно

ИИ-ассистент, который интегрируется в бизнес-процессы клиники и обеспечивает знаниями по запросу.

**Профильная шпаргалка**, которая всегда под рукой.

В виде чат-бота:

- в CRM клиники
- вкладка в браузере
- приложение для ОС

## Реализации чат-бота с ИИ

1. Поддержка сотрудников ресепшена и колл-центра: быстрые ответы на любые вопросы + компиляция персональных предложений для клиентов.
2. Поддержка медицинского персонала процедурных кабинетов.
3. Аналитическая поддержка руководства клиники.

**Система одного окна:** все полезное ищется в одном месте и формулируется в привычном формате (в CRM или в браузере).

## Побочные улучшения

- Сокращение времени обучения персонала.
- Повышение осведомленности персонала о том, что есть в клинике.
- Снижение стресса персонала за счет прямого доступа к гарантированным знаниям.
- Сокращение “дергания коллег по простым вопросам”.



## Оцифрованная выгода

- Сохранение зарплат сервисного персонала **на среднерыночном уровне или ниже.**
- Увеличение возврата клиентов на **8%** за первый месяц работы.

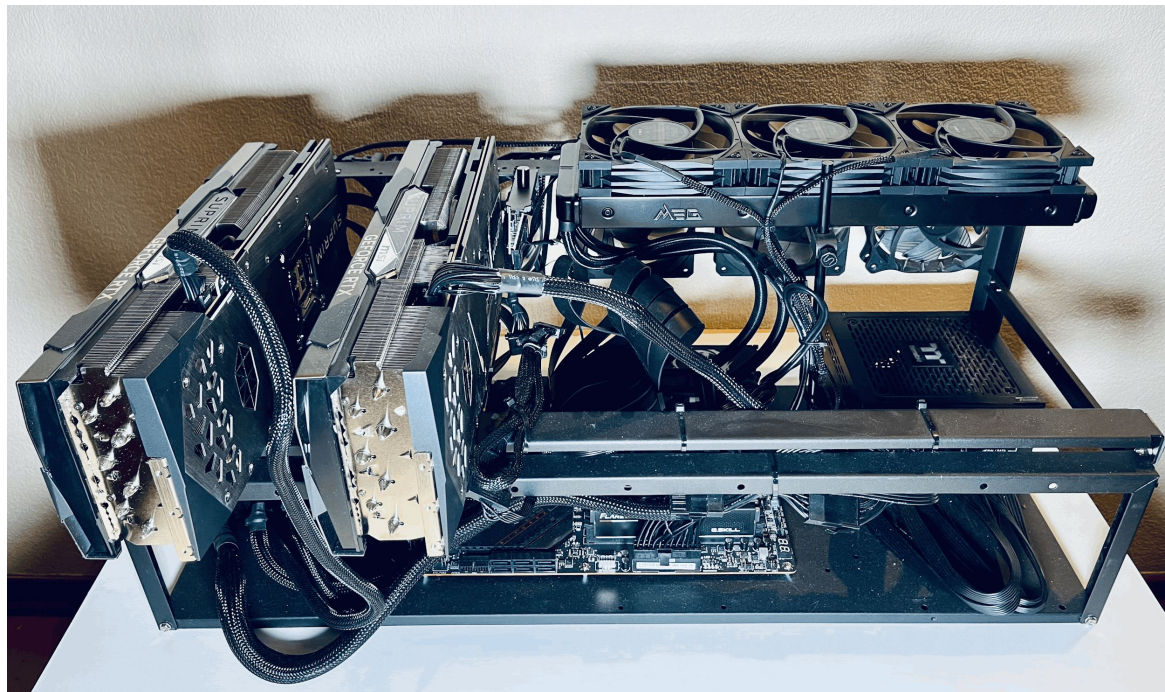
## В чем отличие ассистента с ИИ от корпоративной базы знаний?

- ИИ симулирует поведение “грамотного коллеги”.
- ИИ понимает неточные запросы и осуществляет семантический поиск ответов.
- ИИ формулирует ответ наиболее экономным образом: остается лишь прочитать.
- Сокращает время поиска и время на продумывание запроса.
- Агрегирует последовательный ответ из нескольких баз данных и документов, если в этом есть необходимость.
- Помнит контекст и историю запросов.
- Ориентируется в месте и времени.
- Может искать в интернете, если в этом есть необходимость.
- Проводит анализ данных по поведению пациентов и персонала.

## Техническая реализация

Всего 2 x GPU NVIDIA RTX 4090

Позволяют использовать LLM, достаточные для формулирования быстрых и консистентных ответов.



## Локально, на собственном сервере клиники *on-premise*

- **Отсутствие подписки:** система устанавливается на локальном сервере.
- **Техническая поддержка:** требуется регулярная поддержка.
- **Управляемость:** система способна обрабатывать нагрузку, ограниченную лишь мощностью железа, при этом минимальные требования включают 3 x GPU NVIDIA 4090.
- **Высокая безопасность:** данные остаются внутри клиники.

## Направление развития



Частным медицинским компаниям интересны дешевые ИИ - решения, безопасные с точки зрения защиты чувствительных данных.



1. Чат - бот, вопросы задаются вручную, используется облачный ИИ.
2. То же самое, но полностью локально.
3. Прослушивание разговора и подсказки на лету.
4. Самостоятельные действия агента под присмотром оператора.
5. Агент анализирует сферу, в которой он работает.

## Почему это сложнее, чем кажется?

Какая есть статистика по уровню твоих галлюцинаций?



Галлюцинации в контексте искусственного интеллекта, особенно в языковых моделях, таких как ChatGPT, относятся к генерации уверенных, но неверных или вымышленных ответов. Исследования показывают, что ChatGPT предоставляет правильные или частично правильные ответы примерно в 50,6% случаев, однако предложенные им ссылки существуют только в 14% случаев. [ARXIV](#) Это указывает на значительный уровень галлюцинаций при предоставлении ссылок.

Другие исследования также подчеркивают проблему галлюцинаций в языковых моделях. Например, в одном из обзоров отмечается, что ChatGPT демонстрирует точность около 63,41% в различных категориях рассуждений, что делает его ненадежным в некоторых ситуациях. [ARXIV](#)

Почему это сложнее, чем кажется?



# Галлюцинации

## Ошибки и заблуждения

Тупые баги в фреймворках для работы с LLM

Деградация цепочки рассуждений

- Сложная система самопроверки агента
- Жесткая и разветвленная логика работы
- Длительное тестирование

