ИИ: своё или покупка – затраты и решение

Оцените ИИ-функцию телемедицины по трём слоям затрат, затем пройдите вопросы решения. Суммы – ориентировочные рублёвые оценки, сверяйте с прайсами провайдеров. Это инженерные рекомендации, а не юридическая консультация.

А. Слой 1 · Инференс (видимая вершина)

- [ ] Покупка: лицензия за врача или визит либо цена API за минуту аудио и объём текста
- [ ] Разработка: аренда GPU (ориентир 130–300 ₽ в час) или амортизация своего сервера
- [ ] Пример: визит 15 мин × ≈ 8,5 ₽ за минуту ≈ 130 ₽ за визит
- [ ] 2 000 визитов в месяц × 130 ₽ ≈ 260 тыс. ₽ в месяц «под ключ» за инференс

Б. Слой 2 · Интеграция (чуть ниже ватерлинии)

- [ ] Захват аудио из звонка в функцию
- [ ] Черновик врачу для проверки и подписи УКЭП
- [ ] Запись результата в МИС как черновика на подпись
- [ ] Журнал каждого шага для аудита – нужен и при покупке, и при разработке

В. Слой 3 · Валидация и сопровождение (основная подводная часть)

- [ ] Точность и смещение на вашей популяции до запуска, затем по графику; мониторинг дрейфа
- [ ] Функция только поддерживает врача – или зарегистрирована как медизделие (ПО с ИИ – класс 3)
- [ ] План обновлений: изменения ИИ-ПО-медизделия вносятся по ПП 1684 (п. 112–136)
- [ ] Дежурства, смены версий API; ориентир – senior ≈ 304 тыс. ₽ в месяц на руки

Г. Решение, привязка к поставщику и комплаенс

- [ ] Объём ниже пары тысяч визитов в месяц? → покупка
- [ ] Базовая функция, а не ваше ключевое отличие? → покупка
- [ ] Не потянете сопровождение бессрочно? → покупка
- [ ] Медданные должны остаться в контуре заказчика? → open-модель у себя или on-prem
- [ ] Привязка: уйдёте, если вендор удвоит цену? Изолируйте его за своим интерфейсом
- [ ] Везде: поручение (ст. 6 152-ФЗ), запрет обучения, данные в РФ; зарубежные ИИ-API – нет