



Барьеры внедрения локального GPT и AI- агентов

Гонов Антон Сергеевич



ИЦК ХИМИЯ
И ФАРМАЦЕВТИКА



Почему промышленность внедряет локальный GPT



Защита корпоративных данных



Работа с внутренними знаниями



Интеграция с корпоративными системами



Автоматизация процессов через AI-агентов

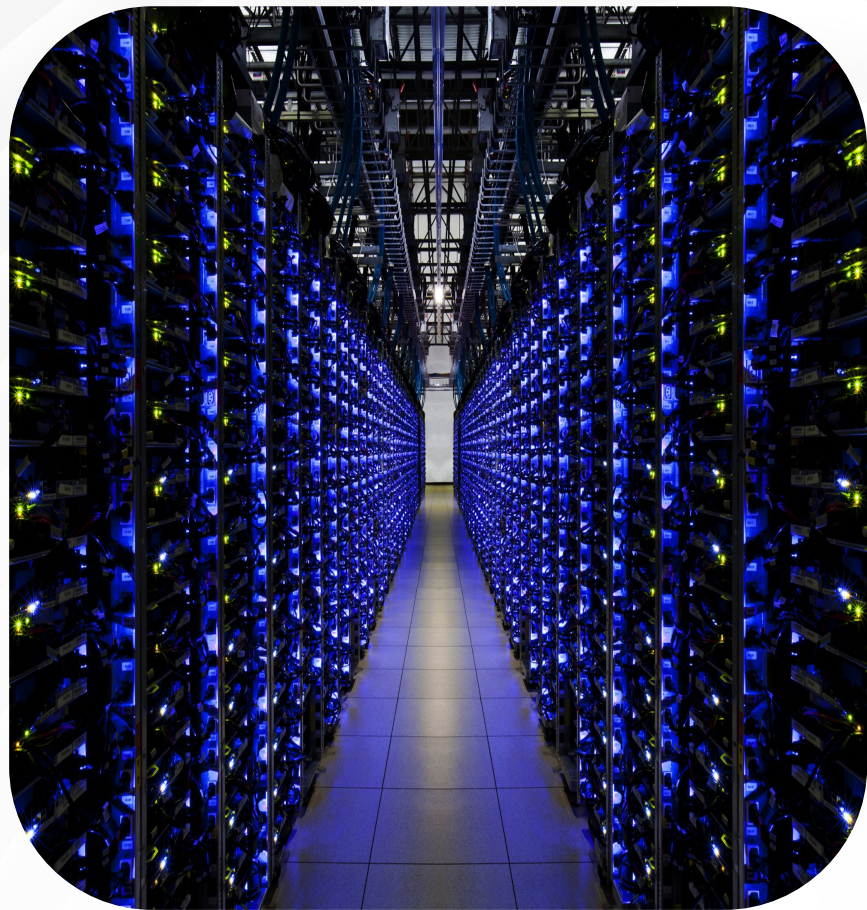
Но внедрение...

Оказывается значительно сложнее, чем ожидалось.



Это не один инструмент, а целая **технологическая экосистема**

Инфраструктура GPU



⚠ Проблемы

Дефицит GPU

Высокая стоимость

Сроки доставки

💰 Типичные цифры

\$250-450K

Сервер 8×H100

\$1-3M

AI-кластер

💡 Решения

Квантование

Стратегическое планирование

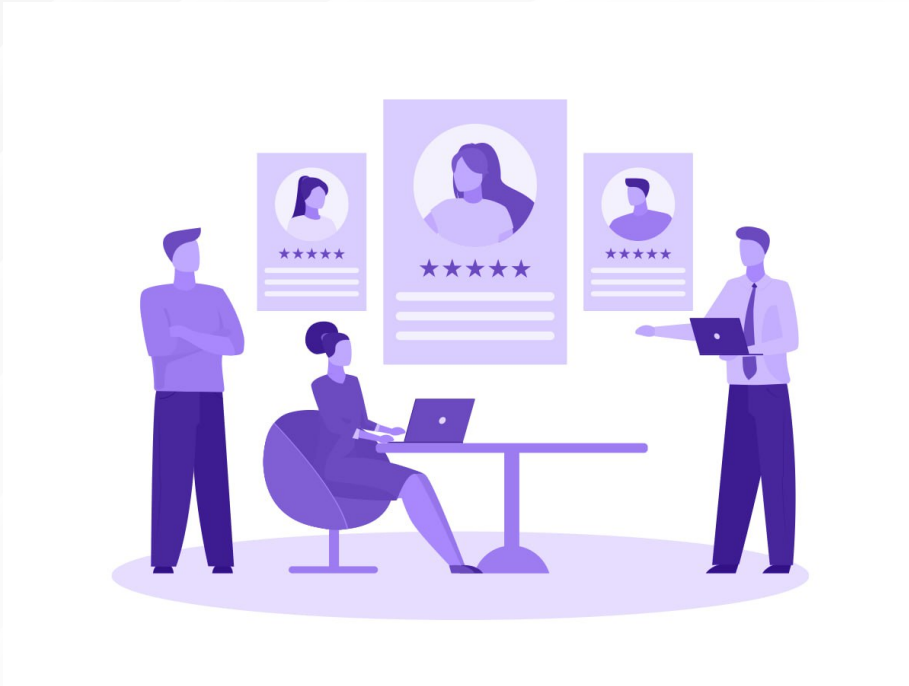
Балансеры

🎓 Урок внедрения

AI-инфраструктуру нужно планировать как отдельную IT-платформу, а не как эксперимент



Кадровый дефицит и компетенции



Новые области

GPU-кластеры

Оптимизация
инференса

Развертывание LLM

Новые процессы

LLMOps

Обучение моделей

Мониторинг качества

Дефицит специалистов

 Специалисты по LLM

 ML-инженеры

 Инженеры AI-платформ

 AI-продакт-менеджеры

Урок внедрения

Внутренние программы обучения • Консультанты на начальном этапе

Пользовательский опыт

← Главная проблема

Пользователь сравнивает внутренний GPT с публичными решениями

ChatGPT

Claude

Gemini



Типичные жалобы

Хуже качество
ответов

Меньше контекста
в диалоге

Медленнее
работает



Решения



Обучать пользователей



Показывать ценность
данных



Постепенно улучшать
модель



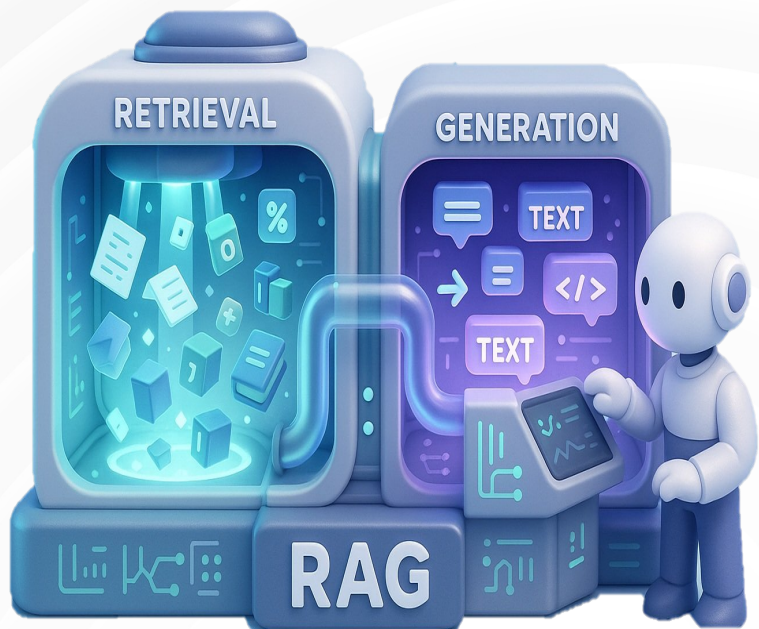
Урок внедрения

Важно управлять ожиданиями пользователей и показывать ценность корпоративных данных





Данные для RAG



LLM без данных компании почти бесполезна



Типичная ситуация



PDF



Сканы



Почта



Папки



SharePoint



Проблемы

Отсутствие структуры

Дублирование

Устаревшие данные



Решения

Стратегия управления данными

Процессы очистки и структурирования

Автоматизированные инструменты



Урок внедрения

70–80% работы в AI-проектах — это подготовка данных и тестирование

Интеграции с корпоративными системами



AI-агенты должны работать с

ERP

CRM

Документооборот

Производственные системы

Основные сложности

Legacy-архитектура
Устаревшие системы

Отсутствие API
Нет интерфейсов

Безопасность
Ограничения ИБ

Решения

Архитектура API-шлюзов

Слой абстракции над legacy

Защищенные каналы интеграции

Урок внедрения

Интеграции часто занимают **больше времени**, чем разработка AI-моделей

Доверие к AI



Главная проблема

Галлюцинации генеративных моделей

Модели могут генерировать достоверно звучащую, но ложную информацию, что подрывает доверие пользователей



Поэтому нужны



RAG



Ссылки на источники



Системы проверки ответов



Ограничение областей применения



Решения

RAG с цитированием источников

Механизмы проверки фактов

Ограничение областей применения



Урок внедрения

Доверие пользователей к AI нужно **специально проектировать**, а не ожидать его по умолчанию



Главный вывод



Это одновременно



Инфраструктура



Данные



Интеграции



Компетенции



Изменение процессов



Ключевой вывод

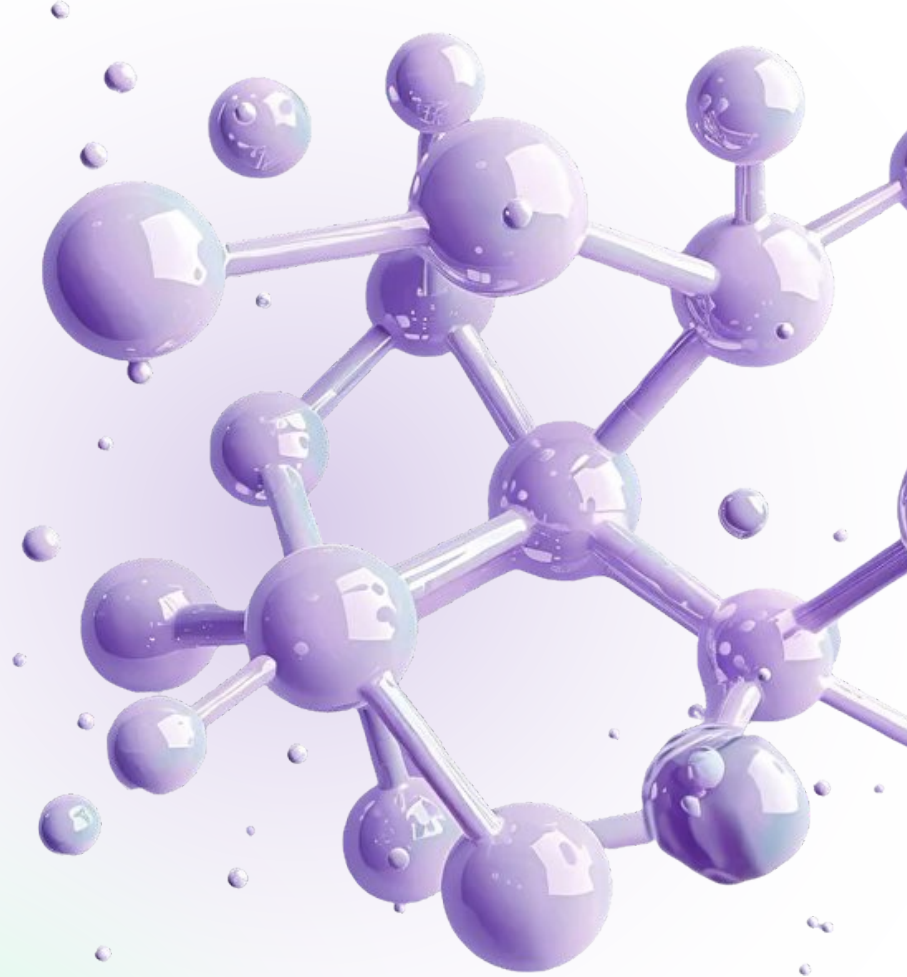
Внедрение AI в промышленности — это не просто внедрение технологии

Это полноценная **цифровая трансформация** компании



Итоговое заключение

Внедрение AI в промышленности — это **полноценная цифровая трансформация компании**, требующая комплексного подхода и системных изменений



Спасибо за внимание!

Гонов А.С.



ИЦК ХИМИЯ
И ФАРМАЦЕВТИКА

