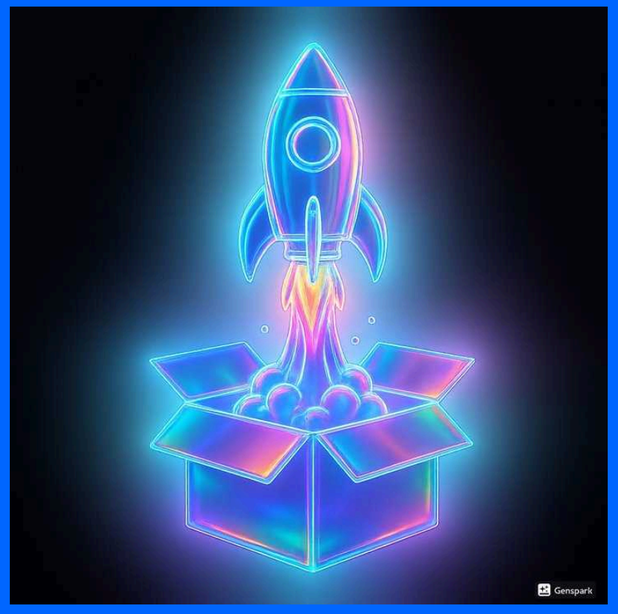
WAPI

Antalya, Turkey | August 2025

ИИ и ИИ-агенты: Зачем нам ИИ — и как им пользоваться без иллюзий?



История, типы, агенты, риски и практика внедрения

План доклада:

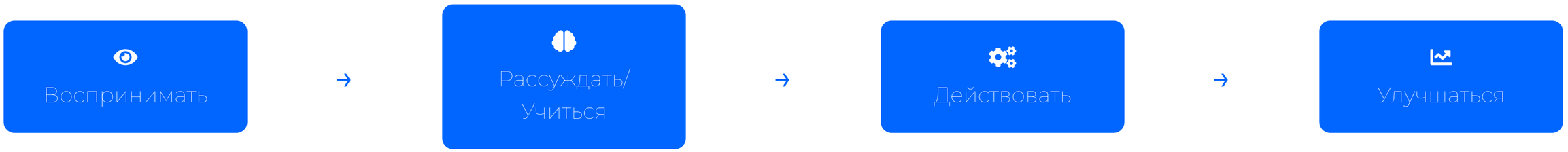
1. Что такое ИИ
2. Почему сейчас
3. Вехи истории
4. Типы/агенты
5. Практика внедрения

1950 — 1975 — 2000 — 2015 — 2025

WAPI

Что такое ИИ: рабочее определение

> ИИ — система, способная **воспринимать** информацию из окружающей среды, **рассуждать и учиться** на её основе, и **действовать** для достижения определённых целей.



Примеры задач:

Распознавание изображений

Понимание текста

Планирование

Диалог

Рациональный агент = выбирает действие, повышающее метрику успеха

Обычная автоматизация

- Действует по заданным правилам
- Не меняет поведение
- Не извлекает новых знаний

Искусственный интеллект

- Адаптируется к ситуации
- Учится на новых данных
- Работает с неопределенностью

WAPI

Почему ИИ «сработал» именно сейчас

3 столпа современного ИИ

Алгоритмы

Архитектура Transformer (2017) революционизировала обработку текста и мультимодальных данных

Прогресс: Внимание вместо рекуррентности, параллельное обучение

Данные

Взрывной рост объемов структурированных и неструктурированных данных для обучения моделей

Масштаб: Триллионы токенов для обучения современных LLM

Вычисления

Рост доступности GPU и облачных вычислений для обучения и инференса нейросетей

Рост: В 100+ раз за последние 10 лет (FLOPS/$)

3 эффекта для бизнеса

Скорость

Резкое сокращение времени на создание прототипов и решение типовых задач

До/После: Часы → минуты на создание текстового контента

Универсальный интерфейс

Естественный язык как универсальный инструмент взаимодействия с ИИ-системами

Снижение: Порога входа для неспециалистов

Снижение издержек

Автоматизация рутинных задач, требовавших высокой квалификации

Экономия: До 30-40% времени в аналитике и разработке

WAPI

История ИИ: ключевые вехи

1950 — 1956 — 1966 — 1997 — 2012 — 2016 — 2017 — 2023 — 2025

1950

Тест Тьюринга

Алан Тьюринг предложил метод определения "разумности" машины через её способность вести беседу как человек.

1956

Дартмутская конференция

Официальное рождение ИИ как научной дисциплины. Джон Маккарти впервые использовал термин "искусственный интеллект".

1966

ELIZA

Первая программа-собеседник, имитирующая диалог с психотерапевтом. Создана Джозефом Вейценбаумом в MIT.

1997

Deep Blue

Компьютер IBM побеждает чемпиона мира по шахматам Гарри Каспарова, демонстрируя возможности ИИ в сложных задачах.

2012

AlexNet

Прорыв в компьютерном зрении с использованием глубоких сверточных нейросетей. Победа в конкурсе ImageNet с большим отрывом.

2016

AlphaGo

ИИ от DeepMind побеждает чемпиона мира по го, древнейшей и сложнейшей стратегической игре.

2017

Transformer

Новая архитектура нейросети "Attention is All You Need" от Google революционизирует обработку естественного языка.

2023

Эра GenAI и LLM-агентов

ChatGPT и другие крупные языковые модели становятся основой для создания автономных ИИ-агентов, меняющих отрасли.

** История ИИ развивается волнами — от теоретических основ к практическим прорывам*

WAPI

Какие ИИ бывают: карта понятий

По возможностям

ANI (Узкий ИИ)

Специализированные системы для решения конкретных задач. Не обладают общим интеллектом.

Пример: распознавание речи, игра в шахматы, рекомендательные системы

AGI (Общий ИИ)

Гипотетические системы с интеллектом человеческого уровня во всех областях.

Статус: концепция, активное исследование, не достигнут

ASI (Сверхинтеллект)

Гипотетические системы, значительно превосходящие человеческий интеллект.

Статус: теоретическая концепция, дискуссия о возможности

По методам

Символический ИИ

Логические правила и представление знаний в виде символов.

Экспертные системы, логическое программирование

ML (машинное обучение)

Алгоритмы, обучающиеся на данных (supervised, unsupervised, RL).

Классификация, регрессия, кластеризация

Глубокие сети

Многослойные нейронные сети для сложных представлений данных.

CNN, RNN, трансформеры

Генеративные модели

Создание нового контента по образцу обучающих данных.

LLM, диффузионные модели, GAN

По задачам

CV (компьютерное зрение)

Анализ и понимание визуального контента.

NLP (обработка языка)

Понимание и генерация естественного языка.

Речь

Распознавание и синтез речи, анализ интонаций.

Планирование

Составление и оптимизация последовательности действий.

Мультимодальность

Работа с разными типами данных одновременно.

WAPI

LLM простыми словами

LLM (Large Language Model) — большая нейросеть на архитектуре Transformer, обученная предсказывать следующий токен (часть слова или знак) в тексте на основе контекста.

Сильные стороны

- Генерация и суммаризация текста
- Написание и рефакторинг кода
- Перевод между языками
- Ведение осмысленного диалога
- Универсальный языковой интерфейс

Слабые стороны

- Галлюцинации (правдоподобная ложь)
- Чувствительность к формулировке запроса
- Ограничения контекстного окна
- Отсутствие актуальных знаний после обучения
- Слабая математика и логические ошибки

Ключевые параметры

Температура
Креативность vs точность
0.0-1.0

Длина ответа
Лимиты токенов для контекста и ответа

Системный промпт
Определяет роль и поведение модели

Пример улучшения промпта

Сырой промпт

"Напиши текст о преимуществах ИИ."

Проблемы: неконкретно, нет аудитории, формата, объема

Улучшенный промпт

"Напиши статью для бизнес-блога о 3 ключевых преимуществах внедрения ИИ в клиентский сервис. Текст на 300 слов с подзаголовками и примером для каждого преимущества."

Улучшено: конкретика, аудитория, формат, объем, структура

WAPI

Что такое ИИ-агенты?

> ИИ-агент — система, которая не только обрабатывает информацию и отвечает на запросы, но **планирует** и **самостоятельно выполняет действия** для достижения поставленных целей.

Цикл работы агента:

Воспринимает → Планирует → Действует → Учится

Уровни автономии:

Советник (Copilot)

Предлагает решения, но финальные действия выполняет человек

Частичная автономия

Самостоятельно решает рутинные задачи, сложные передаёт человеку

Высокая автономия

Самостоятельно выполняет комплексные задачи с минимальным контролем

Сравнение агентов и чат-ботов:

Аспект	Чат-бот	ИИ-агент
Цель	Отвечать на вопросы	Завершать задачи
Действия	Только генерация текста	Использование инструментов, API, систем
Память	Краткосрочный контекст	Долгосрочная память, опыт и знания
Инициатива	Реактивная (ждёт запроса)	Проактивная (может начать действие)

WAPI

Классификация агентов по функциям

Бизнес-функции и примеры использования

Сервис/поддержка

Автоматизация обработки запросов и персонализация поддержки клиентов

Use case:

Автоответчик с доступом к базе знаний, классификация тикетов по приоритетам

Продажи/маркетинг

Персонализация предложений, динамическое ценообразование, лидогенерация

Use case:

Подготовка КП под клиента, динамическое изменение цен на e-commerce

HR

Автоматизация рутинного рекрутинга, адаптации и оценки персонала

Use case:

Первичный скрининг резюме, автоматический поиск кандидатов по вакансии

Финансы/операции

Обработка финансовых документов, анализ расходов, оптимизация процессов

Use case:

Сверка реквизитов и накладных, автоматическая обработка счетов

Технические функции и специальные применения

IT/DevOps

Упрощение разработки, диагностика и устранение инцидентов, автоматизация задач

Use case:

Поиск причин инцидентов, генерация скриптов и тестов, документирование кода

Аналитика/R&D

Обработка данных, обнаружение закономерностей, подготовка отчетов

Use case:

Сводки и репорты с цитированием источников, анализ конкурентов

Роботы/мобильность

Автономное управление физическими системами, взаимодействие с окружающей средой

Use case:

Автономное вождение, дроны доставки, складские роботы

WAPI

Витрина успешных примеров агентов

Реальные примеры ИИ-агентов, успешно решающих бизнес-задачи и создающих измеримую ценность:

Автономное вождение

Восприятие окружения, планирование маршрута и выполнение манёвров в режиме реального времени

Безопасность · Эффективность · Комфорт

Юридический помощник

Анализ юридических документов, подготовка черновиков, выявление рисков и поиск прецедентов

Скорость · Точность · Снижение рутины

Консьерж в отеле

Круглосуточные ответы на вопросы гостей, рекомендации, бронирование услуг и управление запросами

Рост NPS · Многоязычность · Разгрузка персонала

E-commerce агент

Персонализированные рекомендации товаров, динамическое ценообразование и автоматическая поддержка

Конверсия · Маржа · Снижение возвратов

Ключевое отличие: Все эти примеры демонстрируют не просто ответы на запросы, а полный цикл действий — от восприятия и анализа до выбора оптимального решения и его реализации.

WAPI

Анатомия ИИ-агента (архитектура)

LLM-ядро (reasoning)

Основной компонент для обработки естественного языка, рассуждений и принятия решений. Отвечает за понимание запросов и генерацию ответов.

Память

Краткосрочная: хранение контекста текущей беседы. Долговременная: векторная БД для хранения знаний, документов, фактов. Политика забывания: алгоритмы отбора важной информации.

Инструменты/Actions

Функции и API для взаимодействия с внешними системами: Поиск | Базы данных | CRM/ERP | Email | RPA | Код

Планирование/оркестрация

Стратегическое планирование последовательности действий, выбор оптимальных шагов для решения задачи, самопроверка результатов, критерии остановки.

Запрос → RAG/Память → План → Инструменты → Результат

Observability (наблюдаемость)

Система мониторинга и контроля работы агента, обеспечивающая прозрачность и безопасность.

- Логи
- Трассировка
- Метрики качества и стоимости

Принцип: "минимально достаточный набор инструментов"

WAPI

Правила создания агентов (чек-лист)

PEAS-метод проектирования

Performance	Environment	Actuators	Sensors
Метрика успеха	Среда действия	Доступные действия	Входные данные

Ограничить задачу и автономию

Определить четкие границы действий и принятия решений. Начинайте с простых, узких задач.

Human-in-the-loop

Включить контрольные точки для проверки и подтверждения человеком на критических этапах.

RAG вместо дообучения

Использовать Retrieval-Augmented Generation для заземления на факты без дорогого дообучения.

Четкие инструменты и API

Детальные спецификации, валидация входов/выходов, лимиты на время/стоимость/шаги.

Управление памятью

Назначение (что запоминать), срок хранения, политика забывания, приватность данных.

Оценка по метрикам

Офлайн-тестирование, регрессионные тесты, онлайновые KQI (Key Quality Indicators).

Рекомендация:

Начните с пилота на узком сценарии. Планируйте MVP на 2-4 недели с конкретной метрикой успеха и возможностью быстрой итерации.

WAPI

Правила промптов (шаблоны)

Структура эффективного промпта:

Роль → Цель → Контекст/данные → Ограничения → Формат ответа → Критерии качества

Эффективные приёмы:

Few-shot примеры

Дайте 2-3 примера желаемого результата, чтобы модель поняла формат

Step-by-step

Попросите разбить ответ на последовательные шаги

Запрос источников

Попросите указать ссылки на информацию или факты

Стоп-условия

Чётко определите когда нужно остановиться

Анти-паттерны:

- **Расплывчатость**: "Расскажи что-нибудь об ИИ" (без конкретики)
- **Несколько запросов в одном**: смешивание разных задач без приоритетов
- **Отсутствие формата**: не указан желаемый вид результата
- **Противоречивые требования**: несовместимые условия в одном промпте

Мини-кейс: улучшение промпта

Плохой промпт:

"Напиши о применении ИИ в бизнесе"

Улучшенный промпт:

"Выступи в роли бизнес-аналитика. Цель: подготовить краткую справку о 3 ключевых применениях ИИ в логистике для малого бизнеса. Приведи конкретные примеры с ROI, затратами на внедрение и сроками окупаемости. Формат: 3 блока по 100 слов с подзаголовками и списком преимуществ."

Ключевой принцип: чем конкретнее запрос, тем точнее ответ. Сверяйте результат с требованиями и итеративно улучшайте промпт.

WAPI

Проблемы ИИ (честно)

Реальные риски современных ИИ-систем

Галлюцинации

Правдоподобные, но ложные утверждения, которые модель генерирует с высокой уверенностью

> Особо критично в финансовой, юридической и медицинской сферах

Непредсказуемость агентов

Цепочки действий без ограничений могут привести к неожиданным последствиям

> Необходимы "ограждения" и человеческий контроль

Предвзятость

Модели отражают и могут усиливать существующие предубеждения из обучающих данных

> Может проявляться в найме, кредитовании, правосудии

Приватность

Риски утечки персональных данных и несанкционированного доступа к конфиденциальной информации

> Требуется шифрование, анонимизация и строгие политики доступа

Безопасность

Доступ ИИ-агентов к критическим системам может создавать уязвимости

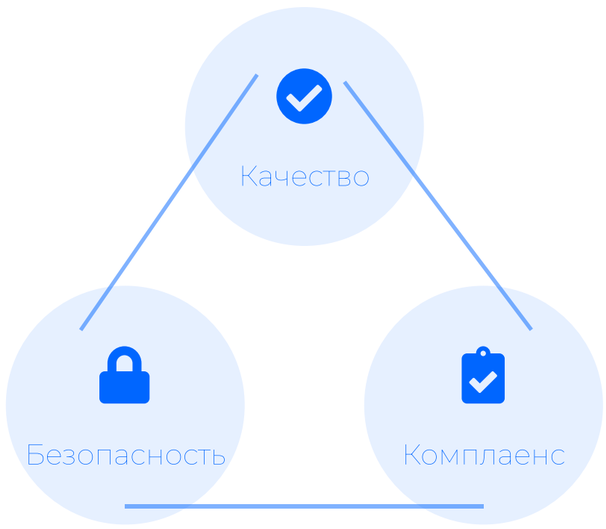
> Необходимы многоуровневые системы авторизации и аудит

Ответственность

Неясность в распределении юридической ответственности за решения, принимаемые ИИ

> Кто отвечает: разработчик, владелец или пользователь?

Треугольник риска: управление критичными аспектами

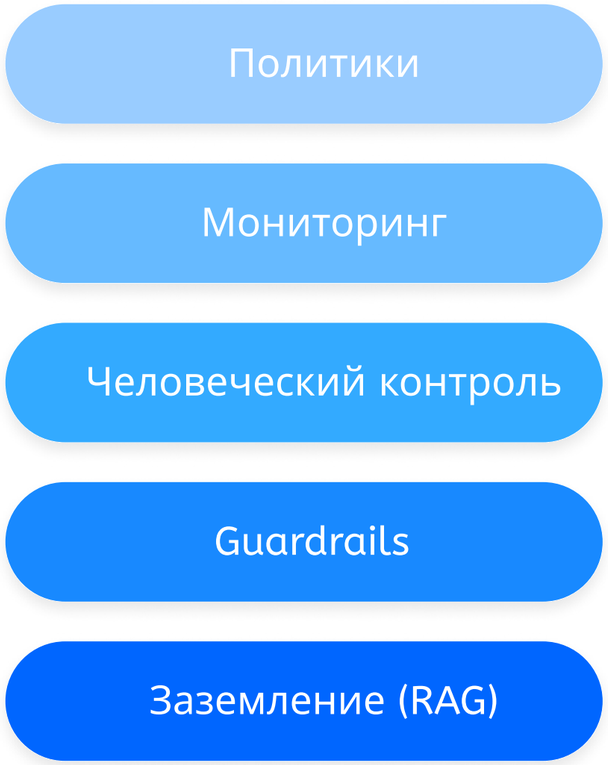


Успешное внедрение требует баланса всех трех факторов.
Пренебрежение любым из них может привести к проблемам в долгосрочной перспективе.

WAPI

Как избегать проблем (практики защиты)

Слоистая защита снижает риски использования ИИ и ИИ-агентов



Заземление через RAG

Retrieval-Augmented Generation обеспечивает доступ к проверенным источникам данных, снижая риск галлюцинаций

Пример: Точность ответов LLM на корпоративные вопросы выросла с 48% до 92% после внедрения RAG с базой знаний компании

Guardrails и ограничения

Whitelist разрешенных инструментов, лимиты на шаги/стоимость, верификация потенциально опасных действий

Человеческий контроль

Одобрение критических действий человеком, проверка предложенных решений, механизмы обратной связи

Мониторинг и логи

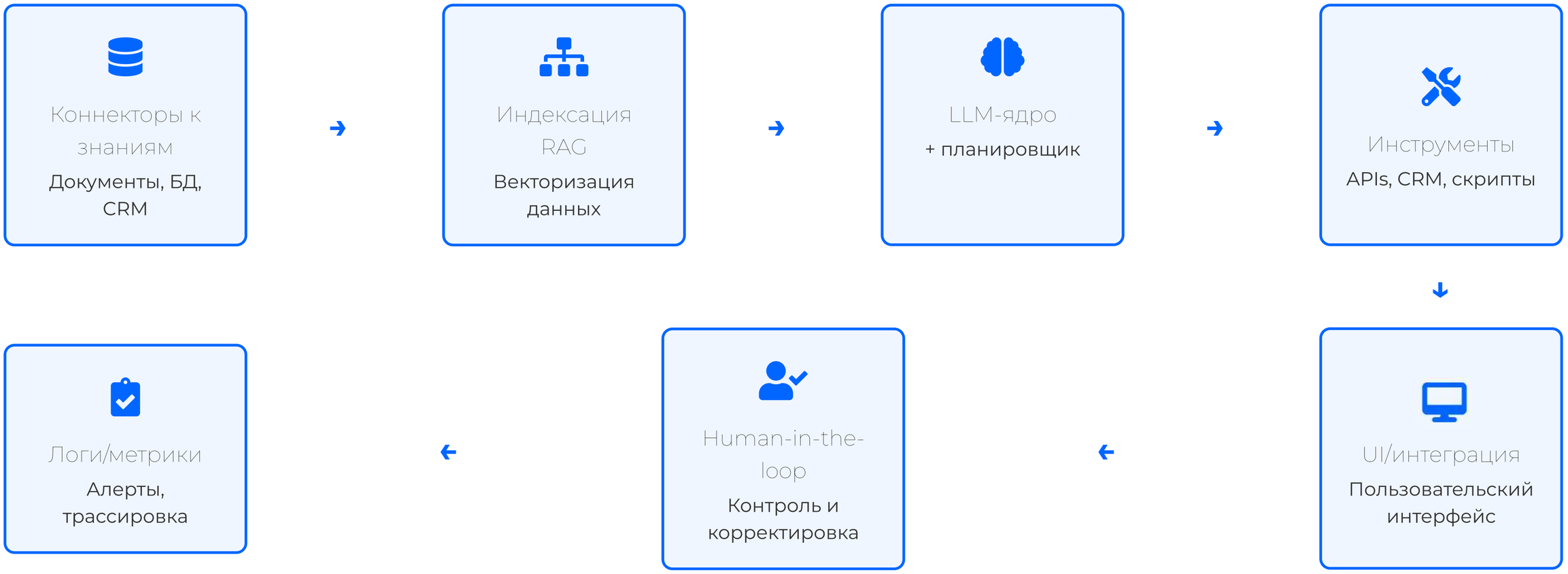
Трассировка цепочек рассуждений, логирование всех действий, алерты об инцидентах, регулярный аудит

Структурированный подход к безопасности

- Определить критичные риски
- Внедрить многослойную защиту
- Регулярно тестировать механизмы

WAPI

Референс-архитектура пилота (2–4 недели)



Узкие места

- Скорость индексации крупных баз данных
- Стоимость запросов к LLM
- Доступ к API и интеграции

Приоритеты MVP

- Ограничить объем знаний (1-2 источника)
- Минимальный набор инструментов
- Фокус на валидации ценности для бизнеса

Типичная продолжительность пилота: 2-4 недели от идеи до тестирования

WAPI

Сферы применения: быстрые победы

Начните с областей, где минимум рисков и максимум ценности

Сервис/Поддержка

Автоответы из базы знаний, автоматическая эскалация, анализ обращений клиентов

Пример:

Агент, отвечающий на 80% типовых вопросов с доступом к продуктовой документации через RAG

Продажи/Маркетинг

Генерация коммерческих предложений, персонализация по ICP, сегментация аудитории

Пример:

Персонализированные e-mail кампании с учетом истории взаимодействия клиента

HR

Создание описаний вакансий, первичный скрининг резюме, составление обучающих материалов

Пример:

Агент для первичного отбора кандидатов, сокращающий время HR на 40%

Финансы/Операции

Сверка документов, проверка реквизитов, автоматизация актов/накладных, выявление аномалий

Пример:

Агент для проверки и сверки счетов с выявлением несоответствий в данных

ИТ/DevOps

Поиск причин инцидентов, генерация кода и скриптов, автоматизация тестирования

Пример:

Диагностика инцидентов и генерация фиксов на основе логов и базы знаний

Аналитика

Создание сводок и отчетов, поиск инсайтов в данных, автоматические аннотации дашбордов

Пример:

Еженедельные текстовые сводки по бизнес-метрикам с цитированием источников

WAPI

Экономический эффект и метрики

Целевые KPI:

Время цикла (TAT)

Сокращение времени выполнения задач

До: 48ч | После: 6ч

Стоимость/задача

Снижение затрат на выполнение операций

До: 1500₽ | После: 300₽

Точность (Acc)

Повышение качества результатов

До: 85% | После: 95%

Формула ROI:

$$\frac{(\text{Сэкономленное время} \times \text{Ставка} \times \text{Объём})}{(\text{Лицензии} + \text{Внедрение} + \text{Инфраструктура})} \times 100\%$$

План измерения эффективности:

1. **A/B тестирование**
   Сравнение работы с ИИ и без на аналогичных задачах
2. **Регулярный мониторинг**
   Непрерывное отслеживание метрик в реальном времени
3. **Оценка бизнес-метрик**
   Конверсия, CSAT, NPS, удержание клиентов
4. **Обратная связь пользователей**
   Качественная оценка удобства использования

Важно: Согласовать KPI с бизнес-владельцами до запуска пилотного проекта

WAPI

Этика и регулирование ИИ (EU AI Act 2025)

Риск-подход

Запрещённый: Манипуляция, социальный скоринг

Высокий: Здравоохранение, транспорт, кредиты

Ограниченный: Чатботы, синтез медиа

Низкий: Спам-фильтры, игры, базовые инструменты

Обязательные практики

Прозрачность
Чёткое информирование пользователей о взаимодействии с ИИ

Документация
Ведение технической документации, инструкций, описаний

Оценка рисков
Регулярный анализ и минимизация потенциального вреда

Управление данными
Контроль качества, обеспечение приватности и репрезентативности

Аудит
Независимая проверка соответствия стандартам и регуляциям

Человеческий надзор
Контроль критичных решений человеком

Роли и ответственность

Владелец системы

Отвечает за правовое соответствие и управление ИИ-системой

Владелец данных

Гарантирует качество и легальность используемых данных

Риск-менеджер

Выявляет и минимизирует риски системы

Комплаенс-офицер

Следит за соблюдением регуляторных требований

Важно: При высоком риске обязательна консультация с юристами и офицером по безопасности данных (DSO)

Тренды 2025–2026: что впереди?

Мульти-агентные системы

2025

Команды специализированных ИИ-агентов, взаимодействующих для решения сложных задач. Каждый агент отвечает за свою область.

Agentic RAG

2025

Эволюция Retrieval Augmented Generation — агенты с многошаговым извлечением, планированием и самопроверкой данных.

LLMOps/наблюдаемость

2025-2026

Инструменты для мониторинга работы ИИ: трейсинг, карточка инцидента, предиктивная аналитика отказов.

Мультимодальность

2026

Слияние текста, изображений, аудио и видео в едином интерфейсе ИИ с продвинутым пониманием контекста.

Бюджетирование токенов

2025

Оптимизация затрат на инференс через умное распределение ресурсов и адаптивные модели разной мощности.

Радар трендов ИИ

Ключевой вывод для бизнеса:

Компании, внедрившие хотя бы 2-3 технологии из радара трендов, получают в среднем на 38% больше ценности от ИИ-инициатив по сравнению с конкурентами.

WAPI

Q&A. Ваши вопросы, следующий шаг!

Вопросы к обсуждению

1. Что автоматизировать первым?
   Какие процессы в вашей компании принесут быстрые победы?

2. Где больше всего теряется времени?
   Какие задачи занимают непропорционально много ресурсов?

3. Какие риски критичны?
   Где ошибки ИИ могут нанести максимальный ущерб?