

ИИ-агенты

для автоматизации задач
бизнеса и государства



ИИ-агенты как ответ на проблемы с качеством работы и дефицитом сотрудников

Среди сотрудников¹

47%

затруднялись найти информацию или данные для выполнения текущей задачи

> 30%

приняли неправильное решение из-за недостаточной осведомленности

> 45%

получили неактуальные данные

36%

постоянно пропускают или не замечают важных обновлений ПО

При этом¹

30%

рабочего дня среднестатистический сотрудник может тратить на поиск информации

> 50%

запросов в поддержку — повторяющиеся вопросы

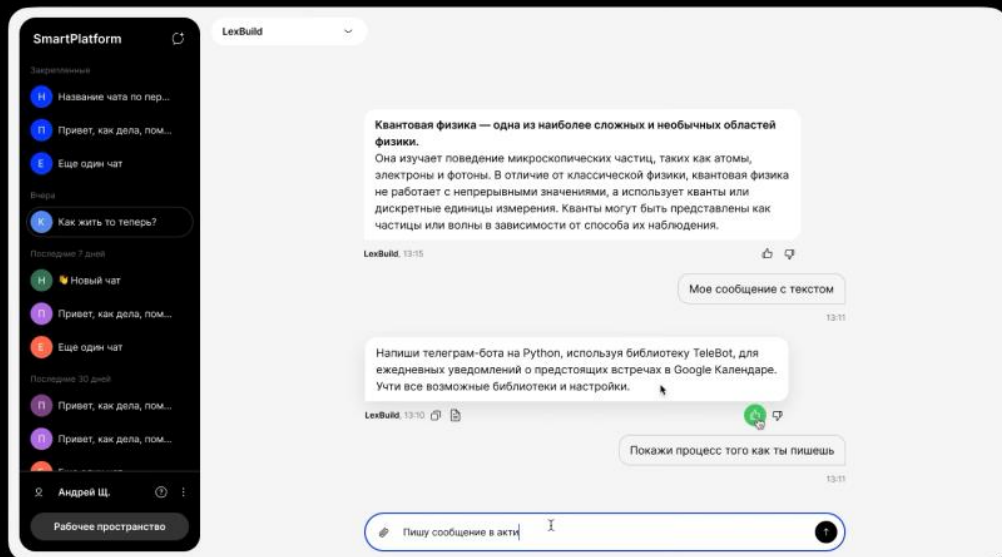
3.1

млн человек
дефицит работников
до 2030 года²

1. по данным IDC, Gartner, McKinsey.
2. о данным Минтруда на январь 2025.

ИИ-агенты¹ — это решение для быстрого поиска, анализа и использования знаний: от поддержки клиентов до аналитики и управления документацией

билайн bigdata&ai | red_mad_robot



до 92%

средняя точность
ответов системы

от 30%

экономия
времени сотрудников



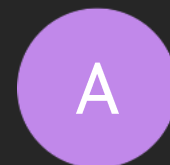
RAG²-модуль

находит релевантную информацию в корпоративных базах, исключая «галлюцинации»



LLM³-ассистент

анализирует и формулирует ответ, подтверждая данные ссылками на источники



Гибкие API⁴

интеграция с экосистемой: CRM⁵, ERP⁶, документами, базами знаний, поддерживает чат-ботов и голосовых ассистентов

1. Обработка данных осуществляется с использованием программного обеспечения. Оценка категорий данных не осуществляется. Клиент гарантирует, что анализ не противоречит законодательству РФ. 2. RAG (Retrieval Augmented Generation) — это технология, сочетающая в себе поиск релевантной информации в существующих хранилищах данных и генерацию текста с помощью языковых моделей для создания более точных и информативных ответов. 3. LLM – большая языковая модель. 4. API – программный интерфейс приложения. 5. CRM – ПО для управления взаимоотношениями с клиентами. 6. ERP – система планирования ресурсов предприятия.

ИИ-агенты¹ на основе корпоративных знаний для автоматизации рутинных задач и оптимизации затрат на ФОТ (фонд оплаты труда)

Билайн bigdata&ai | red_mad_robot

АССИСТЕНТ СЕКРЕТАРЯ

- протоколирование встреч (ВКС)
- саммаризация встреч
- написание резюме

МАРКЕТОЛОГ

- генерация картинок по запросу
- подготовка презентационных материалов

ПОМОЩНИК АНАЛИТИКА

- работа с базами данных (БД): Text-to-SQL², запрос в произвольной форме - выгрузка из БД
- анализ неструктурированной информации

ОПЕРАТОР КОНТАКТ-ЦЕНТРА

- общение с клиентами: работа с претензиями
- создание и актуализация шаблонных ответов
- автоматизация ответов
- отслеживание истории общения с клиентом

ПОМОЩНИК ПРОДАВЦА

- скрипты успешных продаж
- лидогенерация
- кастомизация предложения под клиента
- выявленных паттернов покупательского поведения

и другие

1. Обработка данных осуществляется с использованием программного обеспечения. Оценка категорий данных не осуществляется. Клиент гарантирует, что анализ не противоречит законодательству РФ. 2. Text-to-SQL – технология, которая преобразует запросы на естественном языке в структурированные SQL-запросы.

ИИ-платформа¹ позволяет решать задачи компаний из разных отраслей

Билайн bigdata&ai | red_mad_robot

ФИНАНСОВЫЙ СЕКТОР

- агент по нормативным документам
- агент по финансовой аналитике
- агент по автоматизации клиентской поддержки
- агент по оценке рисков
- агент по сканированию документов различных форматов

РИТЕЙЛ

- агент по обработке клиентских запросов
- агент по логистической аналитике
- агент первой линии поддержки
- помощник менеджера по продажам
- консультант по продуктам

ДЕВЕЛОПМЕНТ

- агент по нормативным документам
- агент клиентского сервиса
- агент по управлению проектной документацией
- агент по аналитике рынка недвижимости

ЛОГИСТИКА И ТРАНСПОРТ

- агент по управлению документацией
- агент по анализу данных поставок
- агент по автоматизации клиентской поддержки
- агент по маршрутизации грузов

МЕДИЦИНА

- агент по анализу медицинских данных
- агент по поддержке диагностики
- агент по обработке запросов пациентов
- агент по регуляторным требованиям

ЛЮБЫЕ ОТРАСЛИ

- агенты поддержки
- помощники/суфлеры
- анализ телефонных/видео звонков
- агенты классификаторы (документация, корреспонденция, почта)

1. Обработка данных осуществляется с использованием программного обеспечения. Оценка категорий данных не осуществляется. Клиент гарантирует, что анализ не противоречит законодательству РФ.

ИИ-агенты¹ работают как единая команда и решают задачи комплексно

1 вы ставите задачу координатору роли

2 координатор подключает нужных агентов

3 агенты выполняют задачи и дообучаются, чтобы давать оптимальный результат

Как это работает на примере:

шаг 1

создать рекламную кампанию по e-mail²

шаг 2

Координатор делит задачу на этапы и подключает нужных агентов. Каждый агент выполняет свою часть задачи



шаг 3

результат = готовое решение

- рекламная кампания
- создана
 - сегментирована
 - оформлена
 - отправлена
 - замерена
 - оптимизирована

Как это будет работать на примере помощника руководителя¹

этапы принятия решения

образ ИИ-агента



уменьшение времени на рутинные задачи

принятие решений на основе данных

уменьшение вероятности ошибки по причине человеческого фактора

¹ Обработка данных осуществляется с использованием программного обеспечения. Оценка категорий данных не осуществляется. Клиент гарантирует, что анализ не противоречит законодательству РФ.

В основе решения — гибридная платформа для создания агентов¹

превращает корпоративные знания в преимущество

✔ больше, чем RAG²

объединяем умный поиск, мультиагентную автоматизацию и адаптацию под бизнес-контекст

✔ интеллектуальные агенты

помощник продавца, оператор контакт-центра, маркетолог, ассистент-секретарь и другие для комплексных бизнес-задач

✔ надежный вариант развертывания

инфраструктура в контуре клиента

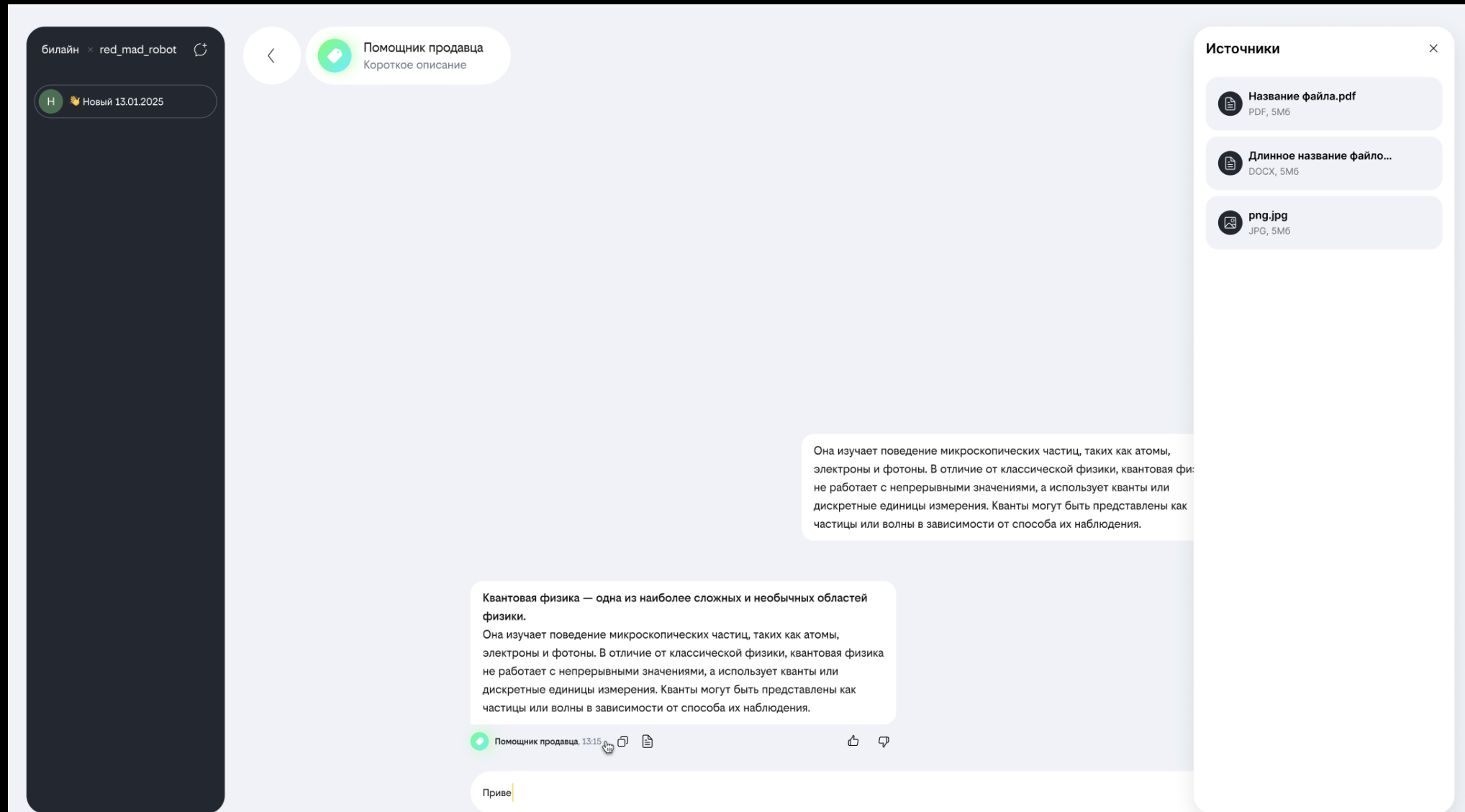
✔ готовое решение с удобным интерфейсом и гибкими API³

работает как самостоятельный продукт или интегрируется в цифровые сервисы

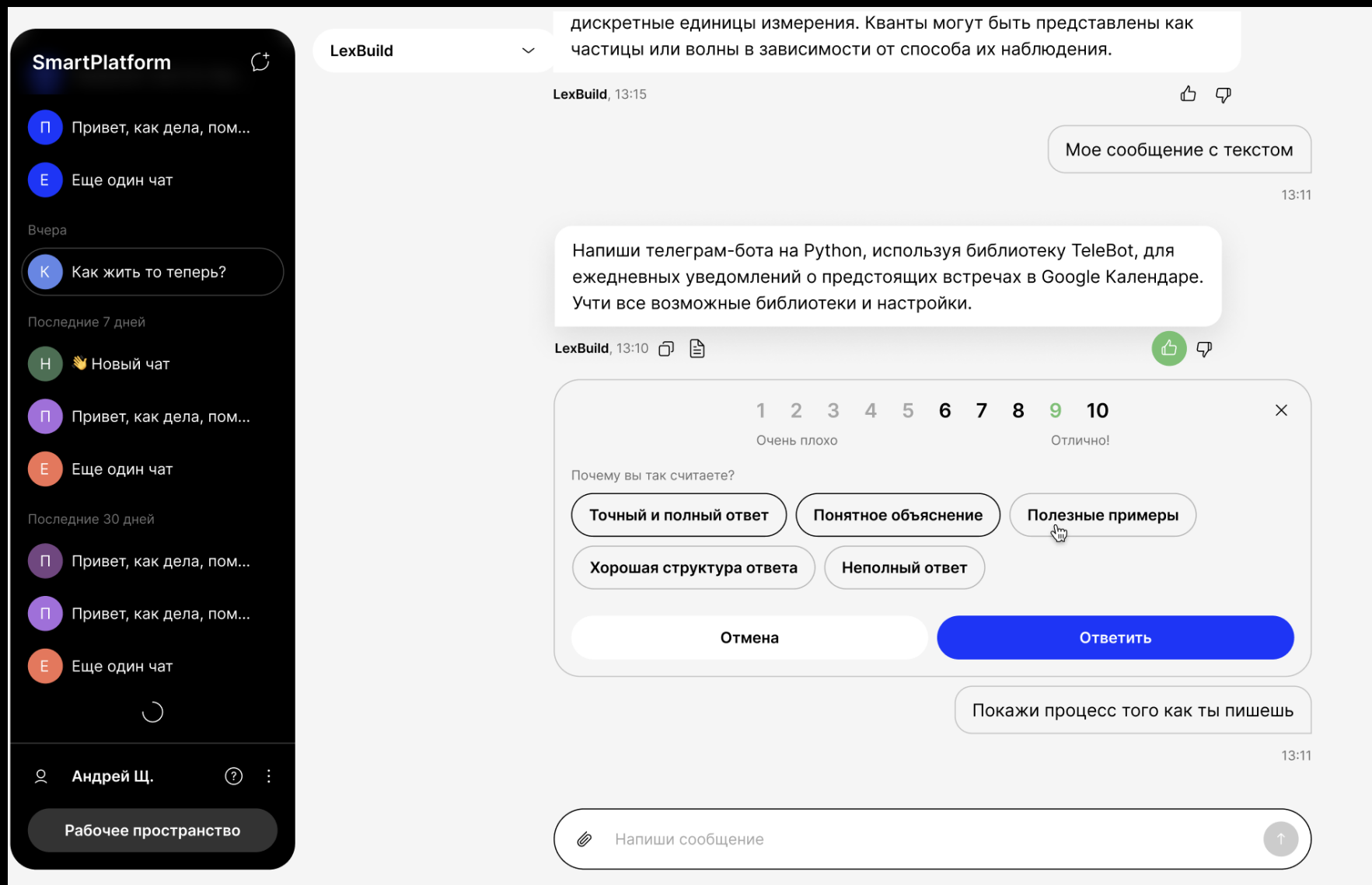
✔ проверено в реальных проектах

до **40%** снижения нагрузки и до **96%** точности ответов⁴

Интерфейсы – поиск по базе знаний



Интерфейсы – генеративный чат



ИИ-платформа усиливает систему управления знаниями, превращая данные компании в ценный и доступный ресурс¹

билайн bigdata&ai | red_mad_robot

ГИБКОСТЬ И КАСТОМИЗАЦИЯ

- индивидуальная адаптация под потребности бизнеса

ЛОКАЛИЗАЦИЯ И БЕЗОПАСНОСТЬ

- развертывание в контуре заказчика
- высокие стандарты безопасности

ПРОЗРАЧНОСТЬ ГЕНЕРАЦИИ ОТВЕТОВ

- ссылки на источники данных для повышения доверия

СКОРОСТЬ ОБРАБОТКИ

- мгновенный доступ к актуальной информации для принятия решений

ЛЕГКОСТЬ ВНЕДРЕНИЯ

- интеграция через API² с существующими бизнес-инструментами

МУЛЬТИКАНАЛЬНОСТЬ

- использование в чат-ботах, голосовых ассистентах и внутренних справочниках

1. Обработка данных осуществляется с использованием программного обеспечения. Оценка категорий данных не осуществляется. Клиент гарантирует, что анализ не противоречит законодательству РФ. 2. API – программный интерфейс приложения.

архитектура и ключевые компоненты

1 →

data processing

коннекторы для сбора данных

- Confluence
- Jira
- Google Drive
- реляционные БД
- кастомные коннекторы

парсинг данных

- text
- images
- audio
- video

разметка данных

- sliding window chunking
- header-based chunking
- contextual and meta chunking
- agents chunking

2 →

agents system

data Markup Agents

- классификация контента
- выделение именованных сущностей
- разметка семантических отношений и создание аннотаций для обучения других ИИ-моделей

Quality Control Agents

- классификация контента, проверка качества разметки, выполненной другими агентами или людьми

Router Agents

- маршрутизация запросов
- анализ входящих задачи, определение приоритетов выполнения

Execution Workflow Agents

- sales Agents
- content Creation Agents
- research Agents
- data Analysis Agents
- support Agents
- и другие

3 →

search engine

Различные поисковые стратегии

- полнотекстовый
- векторный
- гибридный

Методы улучшения качества поиска

- семантическое реранжирование
- HyDE
- Query expansion
- NER

4 →

validation & analytics

Guardrails

- защита от нежелательного контента
- гибкая настройка как базовых фильтров, так и кастомных под нужды конкретной задачи

RAGAS

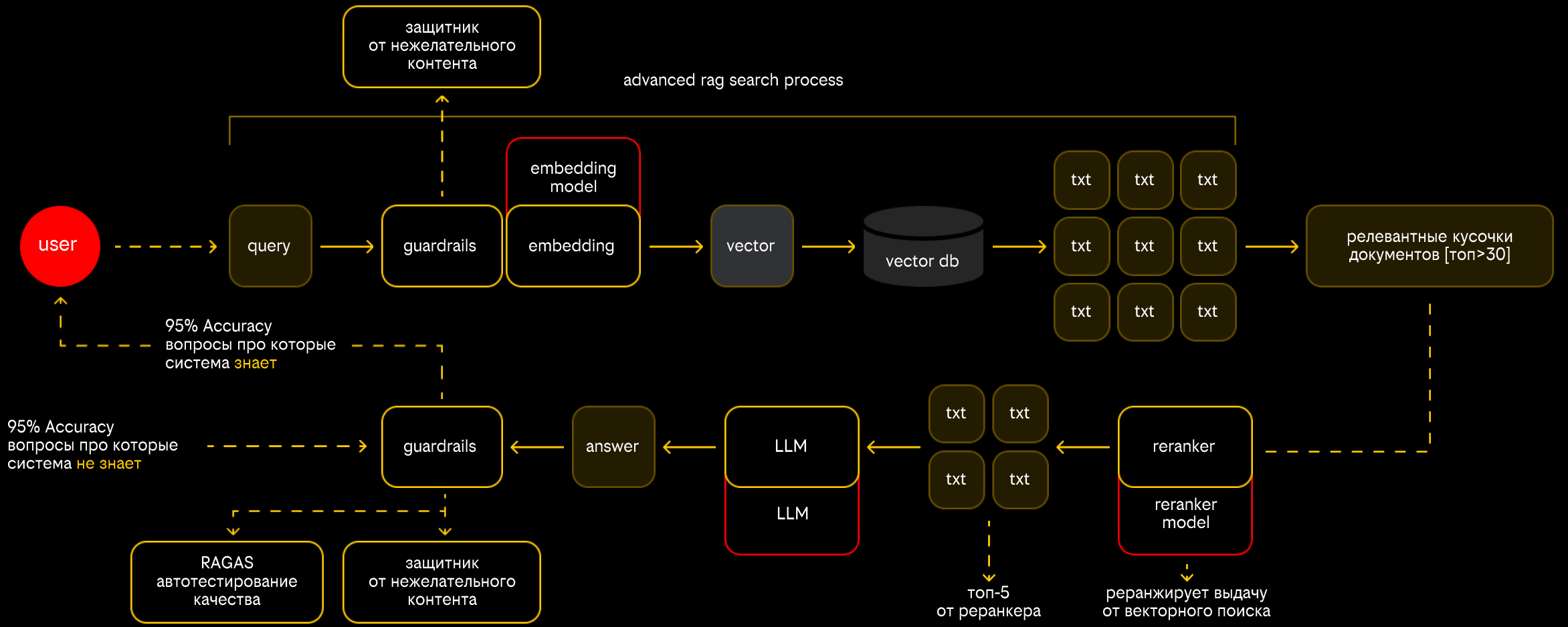
- оценка метрик качества, такие как контекстуальность, согласованность и фактическую точность
- мониторинг качества ответов, что позволяет своевременно обнаруживать деградацию производительности и принимать корректирующие меры

BI

- мониторинг взаимодействия с системой

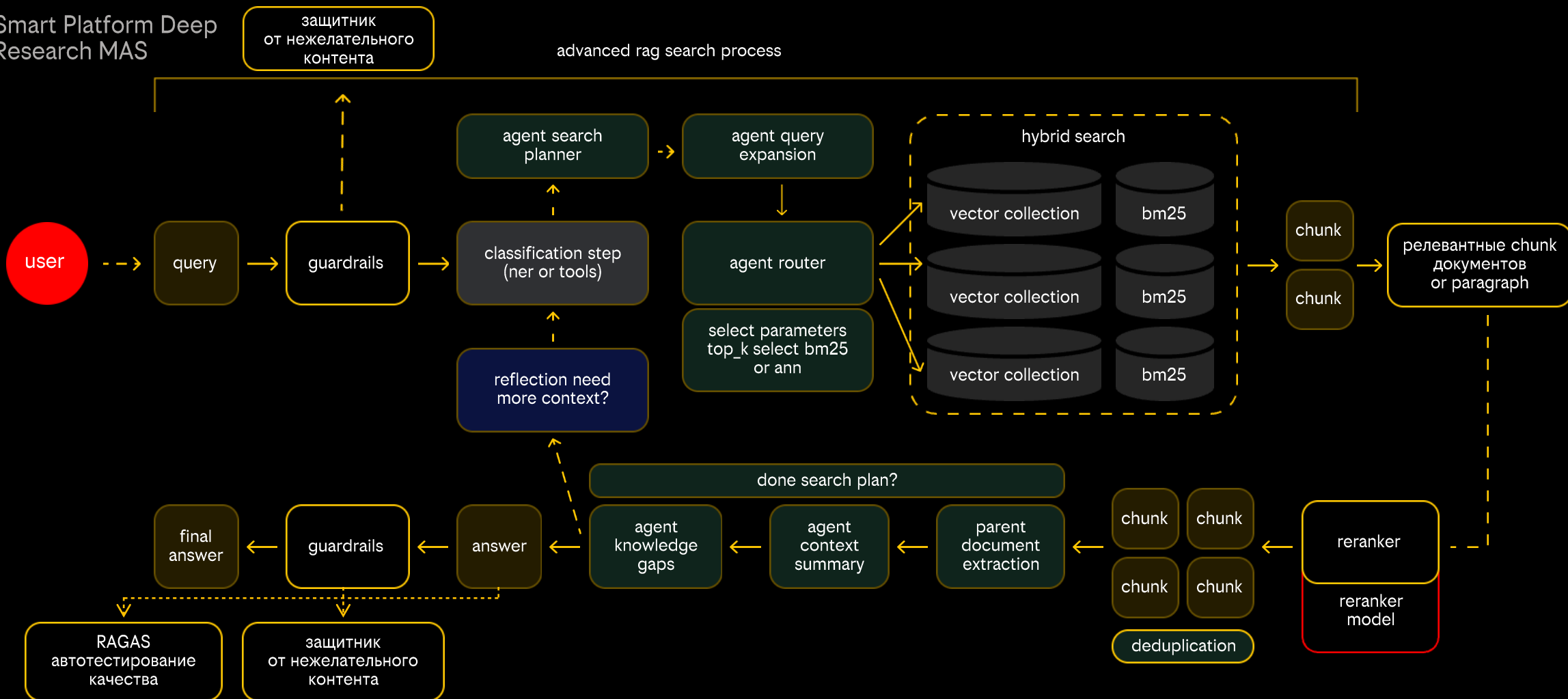
Классический RAG работает в рамках простой цепочки и ограничен в гибкости и адаптации

базовый RAG паплайн



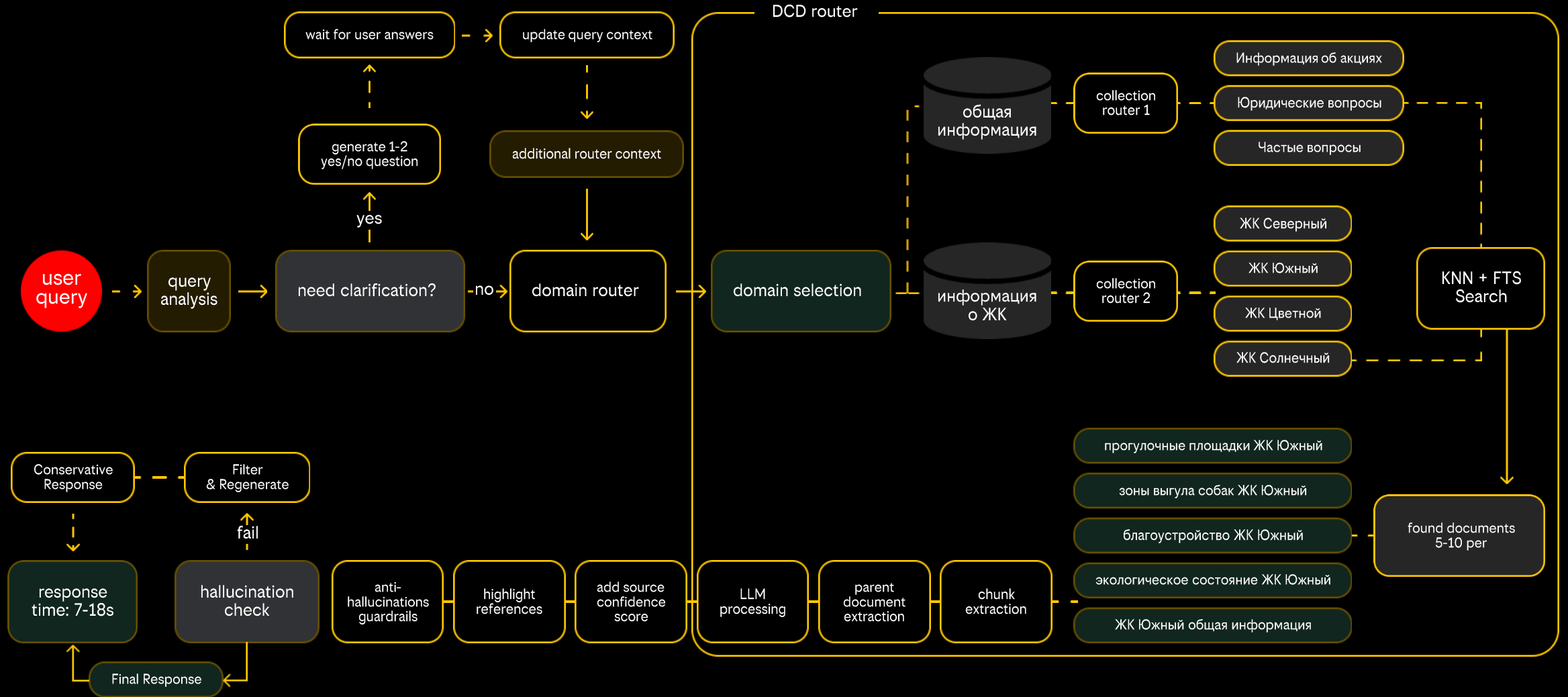
SmartPlatform¹ использует продвинутую агентную архитектуру, чтобы масштабировать RAG под реальные бизнес-сценарии

Smart Platform Deep Research MAS



1. Обработка данных осуществляется с использованием программного обеспечения. Оценка категорий данных не осуществляется. Клиент гарантирует, что анализ не противоречит законодательству РФ.

DCD Архитектура

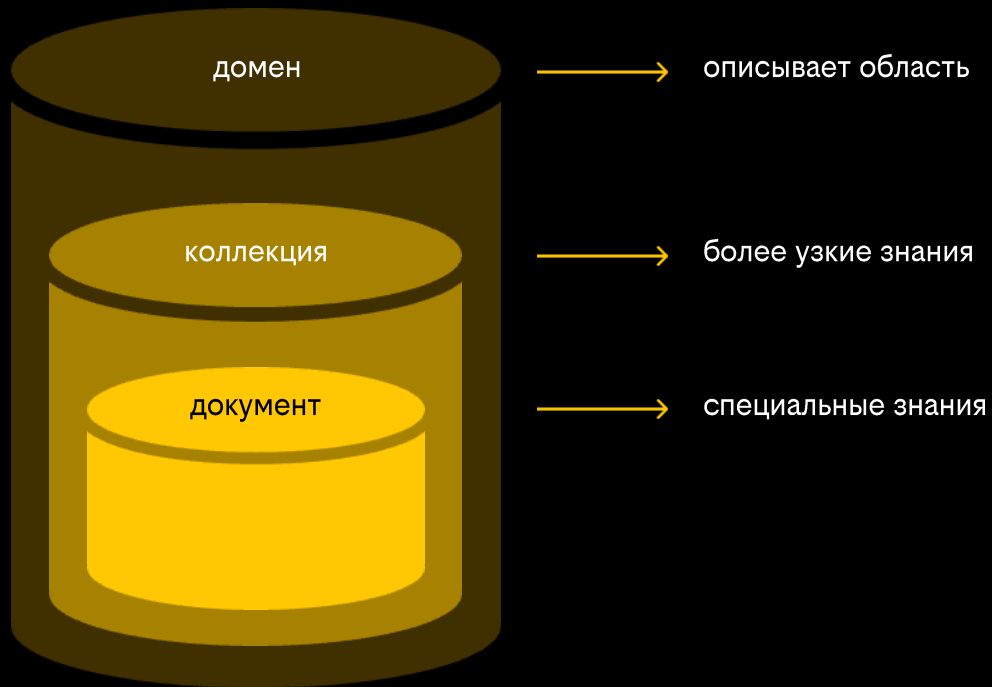


домен > коллекция > документы (DCD design)

билайн bigdata&ai | red_mad_robot

архитектура и компоненты

RAG¹



**доменное разделение знаний или единый домен:
как архитектура ии-системы влияет
на бизнес-результат**

- в ранней версии системы вся информация о жилых комплексах была объединена в единый домен знаний, такой подход казался логичным — один продукт, один набор данных
- на практике он вызвал проблемы: система начала путаться в деталях, информация о разных жилищных комплексах смешивалась, снижалась точность ответов, возникали ошибки в ценах и сроках сдачи

1. RAG – генерация с дополненной выборкой.

кейсы

кейс | ии-ассистент¹ для контакт-центра

билайн bigdata&ai | red_mad_robot

заказчик: билайн, ии-помощник для сервиса «Суфлер»



задача

в момент общения с клиентом давать оператору подсказки, подкрепленные документами из базы знаний (организация умного поиска)

решение:

- конфигурация RAG² на базе знаний заказчика
- разработка интерфейсов – веб-интерфейс и телеграмм-бот
- настройка логирования истории обработки вопросов/ответов
- разработка методики оценки качества работы RAG²-системы

результат³

80-85% правильных ответов на вопросы

получена положительная обратная связь

1. Обработка данных осуществляется с использованием программного обеспечения. Оценка категорий данных не осуществляется. Клиент гарантирует, что анализ не противоречит законодательству РФ. 2. RAG (Retrieval Augmented Generation) — это технология, сочетающая в себе поиск релевантной информации в существующих хранилищах данных и генерацию текста с помощью языковых моделей для создания более точных и информативных ответов. 3. Сведения носят информационно-справочный характер, билайн не гарантирует достижения указанных результатов.

кейс | ии-ассистент¹ продаж

блайн bigdata&ai | red_mad_robot

заказчик: Крупная девелоперская компания

задача

оптимизировать работу отдела продаж, сократить время на поиск информации и подготовку предложений, повысить точность ответов клиентам. главная цель — ускорить сделки, снизить нагрузку на сотрудников и повысить конверсию

решение

запустили ИИ-ассистента для менеджеров по продажам, который ускоряет работу с клиентами и подрядчиками за счёт быстрого поиска информации. он объединяет данные по жилым комплексам в единую базу, помогая менеджерам оперативно отвечать на запросы. использует RAG² - подход для минимизации ошибок, а разметка данных проходит через RAGAS³ для повышения точности

Результат⁴

96%

точность ответов

-40%

снижение нагрузки на отделы

3,7 сек.

среднее время ответа



1. Обработка данных осуществляется с использованием программного обеспечения. Оценка категорий данных не осуществляется. Клиент гарантирует, что анализ не противоречит законодательству РФ. 2. RAG (Retrieval Augmented Generation) — это технология, сочетающая в себе поиск релевантной информации в существующих хранилищах данных и генерацию текста с помощью языковых моделей для создания более точных и информативных ответов. 3. RAGAS — фреймворк для автоматической оценки RAG-систем. 4. Сведения носят информационно-справочный характер, блайн не гарантирует достижения указанных результатов.

кейс | ии-ассистент¹ для контрагентов крупной девелоперской компании

билайн bigdata&ai | red_mad_robot

задача

найти решение, которое сократит время обработки запросов, обеспечит прозрачность сделок и упростит взаимодействие с контрагентами. главная цель — ускорить процессы и освободить ресурсы сотрудников для стратегически важных задач

решение

создали цифрового ассистента на базе генеративного ИИ с применением RAG²-подхода, который объединяет данные из различных источников в единую базу знаний

результат³

+15%

рост конверсии заявок

-20%

снижение нагрузки на отделы

3,7 сек.

среднее время ответа

3 месяца

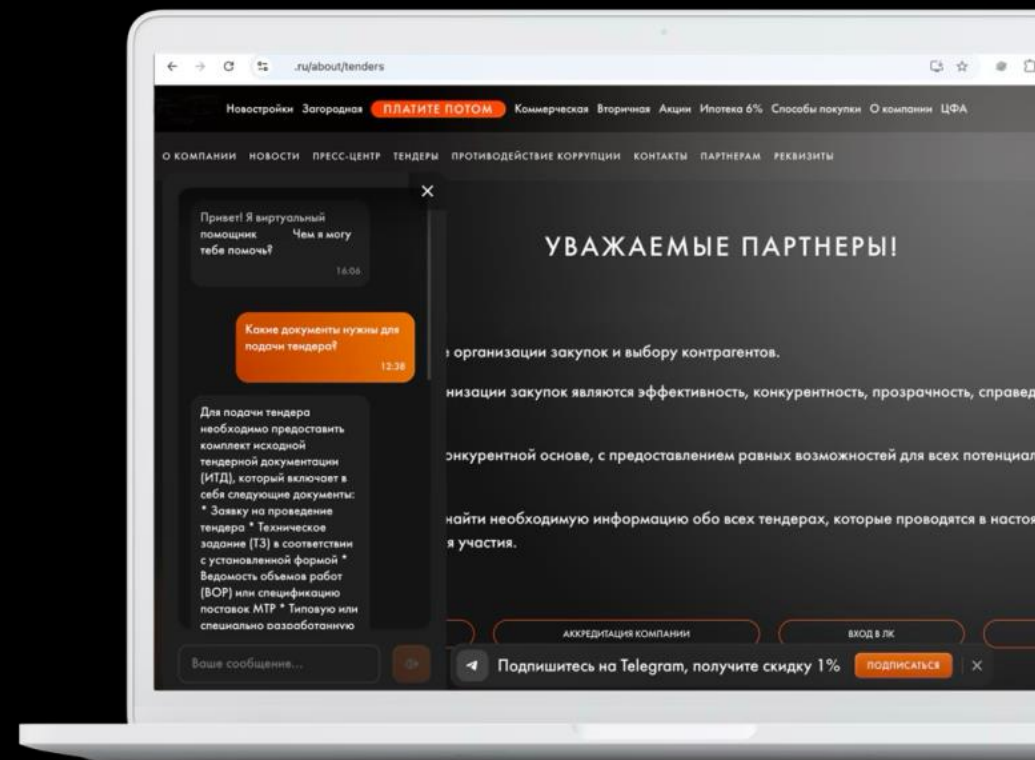
срок окупаемости на проверке гипотез

ROI⁴

550+

экономия человеко-часов в месяц

ROI⁴



1. Обработка данных осуществляется с использованием программного обеспечения. Оценка категорий данных не осуществляется. Клиент гарантирует, что анализ не противоречит законодательству РФ. 2. RAG (Retrieval Augmented Generation) — это технология, сочетающая в себе поиск релевантной информации в существующих хранилищах данных и генерацию текста с помощью языковых моделей для создания более точных и информативных ответов. 3. Сведения носят информационно-справочный характер, билайн не гарантирует достижения указанных результатов. 4. ROI — коэффициент возврата инвестиций.

кейс | ии-ассистент¹ для обработки запросов по ЖК и акциям для одного из крупнейших застройщиков Москвы

билайн bigdata&ai | red_mad_robot

задача

найти решение, которое сократит время обработки запросов, обеспечит прозрачность сделок и упростит взаимодействие с контрагентами. главная цель — ускорить процессы и освободить ресурсы сотрудников для стратегически важных задач

решение

внедрено ИИ-решение, которое автоматически извлекает данные из Confluence и хранит их в векторной базе знаний. агент-роутер классифицирует запросы и направляет их в соответствующий домен, обеспечивая релевантные и быстрые ответы пользователям

результат²

- повышение точности и актуальности ответов за счет автоматического обновления данных из Confluence
- снижение нагрузки на сотрудников, которые теперь могут сосредоточиться на более сложных и стратегических задачах
- увеличение удовлетворенности пользователей благодаря быстрому и точному реагированию на запросы

96%

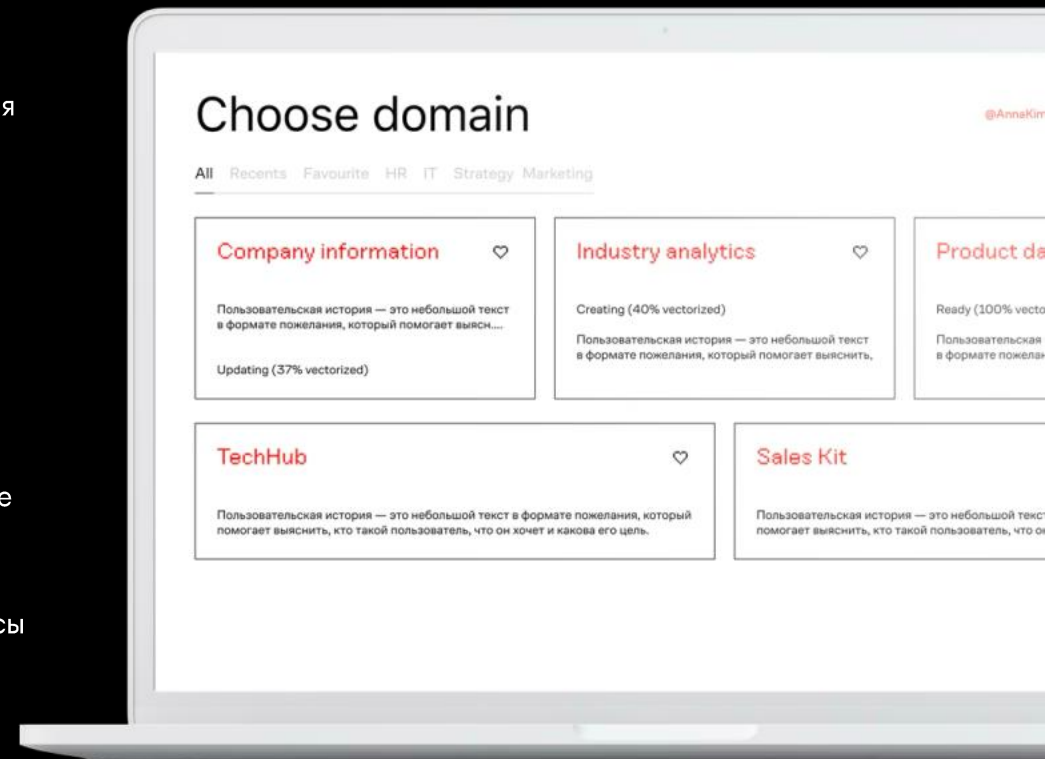
точность ответов

-30%

снижение нагрузки на отделы

3,7 сек.

среднее время ответа



кейс | ии-ассистент¹ для разработчиков корпоративных систем

заказчик: билайн

задача

- ✓ сократить время на типовые задачи разработчиков и автогенерацию тестов и документации
- ✓ улучшить качество кода (авторевью, рефакторинг)
- ✓ обеспечить фильтрацию контента, высокий SLA (соглашение об уровне обслуживания) и масштабируемость



билайн **bigdata&ai** | red_mad_robot

решение:

- внедрена мультязычная Code LLM (RU/EN) в контуре заказчика
- ui+ide-плагин: code completion, автодокументация, тесты
- openai-совместимый api, sso, контекст до 128k токенов
- SLA 99,95%, поддержка и кастомизация моделей ≤3 мес

результат²

20-30%

экономия времени на выполнении рутинных задач

40%

пользователей оценили решение на «хорошо» и «отлично»

улучшено покрытие тестами и повышено качество кода, обеспечен удобный UI/Support-канал

1. Обработка данных осуществляется с использованием программного обеспечения. Оценка категорий данных не осуществляется. Клиент гарантирует, что анализ не противоречит законодательству РФ. 2. Сведения носят информационно-справочный характер, билайн не гарантирует достижения указанных результатов.

кейс | генерация патчей для исправления уязвимостей кода для банка из ТОП-3

задача

помочь разработчикам крупнейшего российского банка сэкономить время на исправление уязвимостей кода, выявленные статическим анализатором кода (SAST¹), повысить качество кода.

решение

разработали мультиагентную систему², которая с помощью LLM³

- обрабатывает срабатывания от SAST-анализатора в формате sarif
- собирает контекст
- классифицирует срабатывания на верные и ложно-положительные
- генерирует исправления кода
- создает merge request с исправлениями
- создает описание уязвимости и исправления, понятное разработчику

разработчик проверяет предложенные исправления и принимает или отклоняет MR⁴

результат⁵

70%
точность классификации

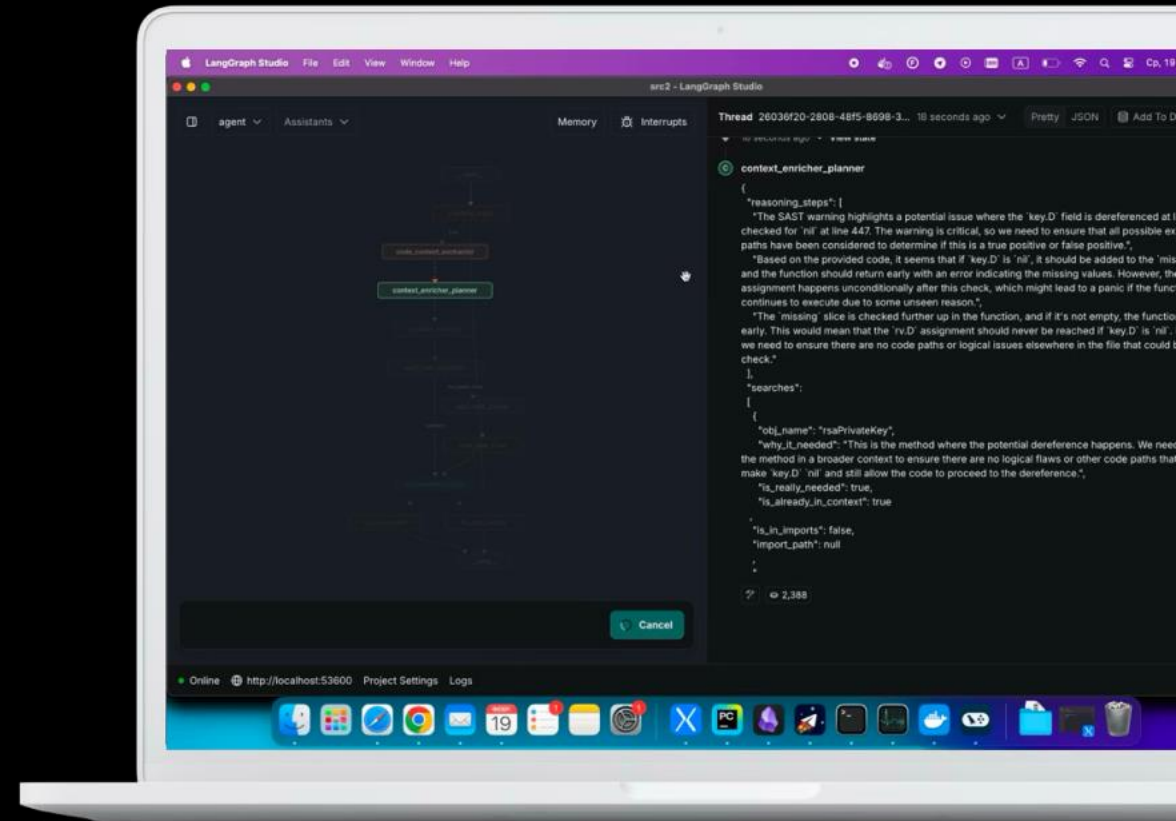
53%
точность исправлений кода

20 минут

повышение качества кода

устранение рутинных задач по описанию патчей

средняя экономия времени на написание одного патча



1. SAST – статический анализ безопасности приложений. 2. Обработка данных осуществляется с использованием программного обеспечения. Оценка категорий данных не осуществляется. Клиент гарантирует, что анализ не противоречит законодательству РФ. 3. LLM – большая языковая модель. 4. MR – запрос на вливание изменений из одной ветки в другую. 5. Сведения носят информационно-справочный характер, билайн не гарантирует достижения указанных результатов.

инфраструктурные требования

необходимая инфраструктура

	базовое решение (до 100 пользователей)	расширенное решение (до 300 пользователей)	корпоративное решение (масштабируемое предприятие)
аппаратная база	2 стандартных ATX-сервера с AMD Ryzen 7	2–4U серверная платформа	4–6U GPU-станция с поддержкой 4–8 GPU
память	64+ ГБ ОЗУ на сервер, 1 ТБ дискового пространства	до 512 ГБ ОЗУ с комбинацией SSD/HDD хранилища, 1 ТБ дискового пространства	от 512 ГБ ОЗУ с комбинацией SSD/HDD хранилища, 2 ТБ дискового пространства
GPU	2× NVIDIA RTX 4090 (по одной на каждый сервер)	NVIDIA A100 + вспомогательная карта (T4/L4/RTX 4000)	Комбинация A100 (для LLM) и H100 (для специализированных задач)
производительность	до 10 параллельных запросов, обработка документов в один поток	до 30 параллельных запросов с 7B моделью	от 100 параллельных запросов
особенности	первый сервер обслуживает основную LLM и базовые сервисы (бэкенд/фронтенд). второй сервер — вспомогательные модели	масштабирование через добавление GPU. возможна установка 32B модели с лучшим качеством диалога (снижает пропускную способность)	нативная поддержка аппаратной квантизации. возможность контекста до 128K токенов. масштабирование: разделение 80 ГБ A100 на 3 инстанса для 7B моделей



VLasovskiy@beeline.ru



ГОТОВЫ ОБСУДИТЬ ваши задачи

Данный материал носит информационный характер и не является офертой. С условиями оказания упомянутых в тексте услуг, тарифов билайн, а также с ограничениями и требованиями к ним, Вы можете ознакомиться на сайте www.beeline.ru в разделе «Бизнесу».