

Корпоративный ИИ-ассистент в закрытом контуре вашей компании

Один вход для всех сотрудников, лучшие языковые модели по одному счёту, а документы и переписка не покидают вашу инфраструктуру.

На вашем сервере

всё в контейнерах, запуск за день

Любые модели

облачные по одному счёту или локальные

Полный контроль

кто, что и на какую сумму использует

Сотрудники уже пользуются ИИ. Вопрос — под чьим контролем

Договоры уходят в чужие сервисы

Люди платят за ChatGPT с личных карт и вставляют туда коммерческие предложения, переписку с клиентами и внутренние документы. Куда это попадает дальше — никто не знает.

ИТ и безопасность ничего не видят

Нет журнала обращений, нет понимания, какие данные и в какие модели отправляются, нет возможности отозвать доступ уволенному.

Расходы размазаны по подпискам

Десятки подписок на разных людях, оплата с карт, компенсации по чекам. Ни бюджета, ни лимитов, ни ответа на вопрос «сколько мы тратим на ИИ».

Рабочее место с ИИ для каждого сотрудника

Привычный чат, к которому подключены лучшие модели, документы компании и ваши внутренние системы.

Чат и модели

Все модели в одном окне. Сотрудник выбирает модель под задачу, а не ищет, где какая подписка.

Документы и знания

Документы загружаются в контур и остаются в нём. Поиск по ним работает на вашем сервере.

Агенты и автоматизация

От ответов на вопросы — к выполнению работы. Агенты подключаются к вашим системам и делают задачи по расписанию.

Подключаемые модули

Подключаются по запросу под конкретный сценарий: каждому нужен отдельный сервис внутри контура или выход наружу, поэтому мы включаем их осознанно.

Чат и модели

Все модели в одном окне. Сотрудник выбирает модель под задачу, а не ищет, где какая подписка.

Несколько моделей в одном окне

Смена модели по ходу беседы, без потери контекста.

Ветвление беседы

Разветвить разговор от любого сообщения и пойти двумя путями.

Библиотека промптов

Общие для компании шаблоны с переменными: «подготовь КП для {клиент}».

Закладки, теги, проекты

Порядок в истории: беседы сгруппированы по задачам и клиентам.

Сравнение ответов

Один вопрос — параллельные ответы двух моделей рядом.

Правка и повторная отправка

Исправить своё сообщение или регенерировать ответ.

Пресеты настроек

Сохранённые наборы: модель, инструкции, параметры — в один клик.

Поиск по всей истории

Полнотекстовый поиск по сообщениям, файлам и фрагментам кода.

Чат и модели

Временные чаты

Беседа без сохранения в историю, когда это нужно.

Импорт и экспорт

Перенести историю из ChatGPT, выгрузить беседу в Markdown, JSON или картинку.

Устойчивая передача ответа

Прерванный ответ восстанавливается сам, без повторного запроса.

Долговременная память

Ассистент помнит предпочтения и контекст сотрудника между беседами.

Обмен беседами

Поделиться перепиской по ссылке — только внутри компании, после входа.

Форматирование

Markdown, формулы LaTeX, подсветка кода, таблицы.

Интерфейс на русском

Тёмная и светлая тема, мобильная версия, горячие клавиши.

Документы и знания

Документы загружаются в контур и остаются в нём. Поиск по ним работает на вашем сервере.

Любые рабочие файлы

PDF, Word, Excel, PowerPoint, текст, JSON, изображения.

Эмбединги внутри контура

Индекс документов считается на вашем сервере — текст никуда не отправляется.

Файл как контекст

Загрузить документ целиком в беседу, когда нужен точный разбор.

Поиск по документам (RAG)

Ответы с опорой на загруженные файлы, с указанием источника.

Общие базы знаний

Регламенты, инструкции, прайс-листы — доступны агентам всей команды.

Документы и знания

Работа с изображениями

Модели со зрением читают скриншоты, схемы, фотографии документов.

Лимиты и типы файлов

Администратор задаёт, что и какого размера можно загружать.

Удалил — значит удалил

Удалённый документ сразу перестаёт находиться поиском.

Хранение на вашем сервере

Никаких сторонних облачных хранилищ для файлов сотрудников.

Агенты и автоматизация

От ответов на вопросы — к выполнению работы. Агенты подключаются к вашим системам и делают задачи по расписанию.

Собственные агенты

Сотрудник собирает помощника: инструкции, файлы, инструменты — и делится им с командой.

Внешние инструменты по MCP

Стандартный протокол подключения инструментов: базы данных, календари, репозитории.

Артефакты

Интерактивные страницы, React-компоненты и диаграммы прямо в чате.

Действия в ваших системах

Вызовы CRM, 1C, ERP, service desk по OpenAPI — агент сам ходит за данными.

Цепочки и подагенты

Сложная задача раскладывается на шаги между специализированными агентами.

Агенты и автоматизация

Личный фоновый агент

Поручение из чата выполняется в фоне; результат — в ленте, когда готов.

Тот же агент в Telegram

Общая память и история: начали в мессенджере — продолжили в рабочем месте.

Задачи по расписанию

«Каждое утро — сводка по заявкам»: агент работает, пока сотрудника нет.

Изоляция и надзор

У каждого сотрудника свой агент, своё рабочее пространство; рискованные действия ждут подтверждения.

Подключаемые модули

Подключаются по запросу под конкретный сценарий: каждому нужен отдельный сервис внутри контура или выход наружу, поэтому мы включаем их осознанно.

Веб-поиск ПО ЗАПРОСУ

Доступ моделей в интернет с переранжированием результатов.

Распознавание текста (OCR) ПО ЗАПРОСУ

Сканы и фотографии документов превращаются в текст для поиска.

Голос ПО ЗАПРОСУ

Распознавание и синтез речи: диктовка и озвучивание ответов.

Интерпретатор кода ПО ЗАПРОСУ

Изолированное выполнение Python, JavaScript, Go, Rust: расчёты, графики, обработка таблиц.

Генерация изображений ПО ЗАПРОСУ

Иллюстрации и макеты по описанию.

Локальные модели ПО ЗАПРОСУ

Собственные модели на вашем GPU для полностью автономной работы.

Управление, учёт и контроль — из одного места

ИТ-служба получает ту же управляемость, что у любой корпоративной системы: кто входит, что ему можно и сколько это стоит.

Корпоративный вход

Единый вход (SSO): подключение Active Directory, LDAP, SAML и других провайдеров. Второй фактор для администраторов. Уволили — доступ пропал везде одним действием.

Учёт расходов и бюджеты

Расходы видны по каждому сотруднику, ключу и модели. Бюджеты и лимиты на ключах и командах, один счёт на всех провайдеров вместо десятков подписок.

Каталог моделей под вашим контролем

Qwen, DeepSeek, GLM — из коробки; OpenAI и Anthropic — по запросу; локальные модели через Ollama или vLLM. Вы решаете, какие модели видит каждая группа.

Ключи провайдеров не покидают шлюз

Сотрудники и агенты никогда не видят ключей OpenAI, Anthropic или другого провайдера. Они получают виртуальные ключи с бюджетом и списком разрешённых моделей.

Журналы и резервные копии

Журнал входов и административных действий. Ежедневные копии базы с настройками доступа и автоматической проверкой, что копия восстанавливается.

Под вашим брендом

Название, логотип и домен компании. Никаких подписей стороннего продукта на экранах сотрудников.

Всё внутри периметра. Наружу — только один шлюз

Платформа целиком работает на вашем сервере. Единственная точка выхода в интернет — шлюз моделей, и её можно закрыть совсем.



Единственный выход наружу — шлюз

Внутренние сети не имеют маршрута в интернет. Любой вызов модели проходит через шлюз, где он учитывается и журналируется.

Каждому сотруднику — свой агент

Фоновые агенты изолированы по контейнерам и рабочим каталогам. Данные одного сотрудника недоступны агенту другого.

Один сервер, один день

Всё поставляется в контейнерах и разворачивается на одном Linux-сервере. Не нужно ни специального оборудования, ни отдельной команды сопровождения.

Закрытый контур — это архитектура, а не настройка

Изоляция заложена в топологию сети, а не в галочку в интерфейсе, которую можно случайно снять.

Данные не покидают сервер

Файлы, история бесед, индекс документов и память агентов хранятся только у вас. Никаких сторонних облачных хранилищ.

Секреты только в шлюзе

Ключи провайдеров существуют в одном месте. Сотрудники и агенты работают по виртуальным ключам с бюджетом и списком моделей.

Жёсткая политика входа

Корпоративный SSO, второй фактор для администраторов, политика паролей, защита от перебора, отсутствие самостоятельной регистрации.

Маскирование персональных данных

Фоновые агенты умеют редактировать персональные данные перед отправкой в модель.

Внутренние сети без выхода в интернет

Рабочее место, базы данных и агенты не имеют маршрута наружу. Единственный выход — через шлюз моделей.

Изоляция сотрудников

У каждого — свой фоновый агент, своё рабочее пространство и свой ключ. Сессии, история и файлы не пересекаются.

Журнал всего

Входы, административные действия, обращения к моделям с разбивкой по людям и ключам. Есть что показать аудиту.

Полностью автономный вариант

Локальные модели на вашем GPU — и контур вовсе не нуждается в интернете.

От демо до рабочей системы — три шага

Демо по запросу

Показываем платформу на нашем сервере и выдаём доступ вашей команде, чтобы попробовать на реальных задачах. Обсуждаем сценарии, модели и требования безопасности.

Пилот на вашем сервере

Разворачиваем контур в вашей инфраструктуре, подключаем корпоративный вход и первую группу из 10 сотрудников. Настраиваем модели, бюджеты и документы.

Масштабирование и передача

Подключаем остальные отделы, интеграции с вашими системами и агентов под ваши процессы. Передаём администрирование вашему ИТ и остаёмся на поддержке.

Что обычно спрашивают

Где хранятся наши данные?

На вашем сервере. Файлы, история бесед, индекс документов, память агентов и настройки доступа лежат в вашей инфраструктуре. Наружу уходит только текст запроса к выбранной модели — через шлюз, где это учитывается. С локальными моделями наружу не уходит ничего.

Какие модели доступны? Можно ли свои?

Из коробки — Qwen, DeepSeek и GLM через единый счёт. OpenAI и Anthropic подключаются по запросу. Локальные модели на вашем GPU — через Ollama или vLLM. Вы решаете, какие модели видит каждая группа сотрудников, и можете отключить любую.

Сколько сотрудников выдерживает один сервер?

Десятки сотрудников работают на одном сервере. Для сотен нужна кластерная установка — мы поможем спроектировать её на этапе масштабирования.

Что обычно спрашивают

Какие требования к серверу?

Linux-сервер с Docker. Для базовой поставки с облачными моделями достаточно нескольких ядер и 16 ГБ памяти; GPU нужен только для локальных моделей. Точные требования зависят от числа сотрудников и сценариев — обсудим на демо.

Чем это отличается от ChatGPT Team или похожих подписок?

Подписка живёт у провайдера: данные хранятся у него, модели — только его, доступ и расходы управляются на его стороне. Locus живёт у вас: любые модели, корпоративный вход, собственные интеграции и агенты, полный контроль над данными и расходами — и один счёт вместо десятков.

Что обычно спрашивают

Как насчёт 152-ФЗ и персональных данных?

Платформа находится в вашем контуре, и вы остаётесь оператором данных. Мы помогаем разделить сценарии: что можно отправлять в облачные модели, что — только в локальные. Агенты умеют маскировать персональные данные перед отправкой в модель.

Можно ли ограничить модели или функции отдельным отделам?

Да. Доступ к моделям и бюджеты задаются на уровне ключей и команд, административные функции — через группы корпоративного входа. Например, юристы работают только с локальной моделью, а маркетинг — с облачными.

ДЕМО

Посмотрите Locus на реальных задачах

Оставьте заявку — мы свяжемся, покажем платформу и выдадим вашей команде доступ к демо-контуре.

30 минут — обзор платформы под ваши сценарии

Доступ для нескольких сотрудников на время пробы

Оценка требований к серверу и плана пилота

Почта **petr@locus-team.ru**

Telegram **t.me/beezzoo**

ООО «БиАй Лабс» Москва