



Разработка **ИИ** решений для бизнеса

Разрабатываем ИИ-решения под вашу задачу — от пилота до внедрения с конкретным результатом

50+

проектов реализовано

2–6

недель до результата

10+

отраслей работы

×3

рост эффективности

Узнаёте себя? Тогда нам есть, о чём поговорить

[d.t]



Хотим ИИ, но не знаем, с чего начать

Есть ощущение, что ИИ может помочь, но непонятно как и что это даст. Не ясно, как удачно внедрить автоматизацию.



Знания есть, но нет системы

Регламенты в папках, инструкции в почте, ответы у конкретных людей. Новый сотрудник учится 2–3 месяца, а с уходом эксперта теряется информация.



Данные есть, а на решения нет времени

Отчёты раз в месяц, аномалии замечают постфактум. Пока собрали данные — момент упущен. Решения принимаются по вчерашней картине.



Пробовали — не получилось

Подключили готовое решение или API — дало нечёткий результат, не встроили в процессы. Разработка ИИ-решения под конкретную задачу — только так оно по-настоящему работает.

В каждом из этих случаев мы начинаем с диагностики и за 48 часов даём конкретный план

Разрабатываем ИИ-решение под вашу задачу с нуля

[d.t]

DZETA

Команда ML-инженеров, которая пишет ИИ под вашу задачу с нуля. Не внедряем готовые коробочные решения, не настраиваем ChatGPT — разрабатываем ИИ-решение с нужной архитектурой, тестами и метриками качества.



Инженерный подход

Проектируем архитектуру под задачу: RAG, классификация, маршрутизация, агенты. Обосновываем необходимость каждого решения.



RAG, не галлюцинации

Системы строятся на ваших данных, а не на «знаниях» общей модели. В каждом ответе ссылка на источник.



Результат до запуска

До начала разработки фиксируем метрики: точность ответа, время ответа. Сдаём код только когда метрики достигнуты.

Команда с опытом в Enterprise

[d.t]

Наши инженеры работали в крупнейших технологических компаниях России до того, как пришли в DZETA. Мы знаем, как устроены промышленные системы изнутри, и строим по тем же стандартам.

Корпоративный опыт



ML-инженеры,
рекомендательные
системы, NLP



НОРНИКЕЛЬ

Промышленный
ML, IoT-аналитика,
аномалии



СБЕР

CV, NLP, разработка
продуктовых ИИ-фич

Академическая база



Шесть типов ИИ-систем, которые мы разрабатываем



Каждая из них — кастомная разработка: своя архитектура, свои данные, свои метрики качества



Умная поддержка

Разрабатываем систему на ваших диалогах: классификация тематики, RAG по базе знаний, эскалация на оператора. 70–80% обращений — без участия людей.



Корпоративный поиск

Регламенты в папках, инструкции в почте, ответы у конкретных людей. Строим понятные системы, не завязанные на конкретных сотрудниках.



Ассистент продаж

ИИ-ассистент, встроенный в CRM: квалифицирует лидов по вашим критериям, генерирует черновики КП, фиксирует следующий шаг.



Обработка документов

Разрабатываем пайплайн извлечения данных: OCR + NLP + валидация. Договоры, счета, акты — структурированные данные без ручного труда.



Аналитика и прогнозы

ML-модели на ваших данных: прогнозирование спроса, детектирование аномалий, сегментация. Не BI-дашборды, именно предиктивная аналитика.



Голосовые ассистенты

Speech-to-text + классификация + резюме. Разрабатываем под вашу телефонию: каждый звонок — тема, тональность, следующий шаг за 30 сек.

Как мы разрабатываем ИИ-решения

[d.t]

01



Диагностика

Разбираем задачу: что должна делать система, на каких данных, как измерить результат. Метрика успеха устанавливается до написания кода.

02



Подготовка данных

Подключаемся к вашим источникам, чистим и размечаем данные. Без этого шага решение не будет работать точно: мы не пропускаем его ради скорости.

03



Пилот за 2–4 недели

Разрабатываем и тестируем 1–2 сценария на ваших данных. Вы видите живую систему с реальными метриками до полноценного внедрения.

04



Интеграции

Встраиваем систему в ваш стек: CRM, 1С, телефония, мессенджеры. Данные остаются в вашей инфраструктуре.

05



Запуск и поддержка

После запуска — мониторинг качества модели. При скачках точности или изменении данных: алерт, переобучение, новый релиз.

Кейс: Ноль пропущенных отказов на авиационном производстве

[d.t]

ML • Детектирование аномалий • Машиностроение • Авиакomпоненты

-40%

времени на
анализ данных

0

пропущенных
критических аномалий

3 мес

срок
выполнения

0 проекте

- ML-модели, обученные на историческом браке — система сама пересчитывает пороги при изменении условий
- Инженеры получают алерт до поломки, а не после: переход от реактивного к превентивному обслуживанию
- Сократили время анализа на 40%, критических пропусков — ноль за 8 месяцев эксплуатации
- Тысячи измерений датчиков в сутки — аномалии замечали только после брака или поломки, мы это исправили

Стек технологий

- Python · scikit-learn
- Apache Kafka
- InfluxDB
- Grafana
- Docker · VPS

Кейс: Отчёт за 15 минут вместо двух дней

[d.t]

Транспортная инфраструктура • Государственный сектор • Автоматизация • Документы

-60% времени на подготовку
отчётности

15 мин вместо 2х
дней

6 нед срок
выполнения

О проекте

- Ежемесячный инфраструктурный отчёт — 2 дня ручного сбора данных из 5 разных систем
- Автоматический сбор через API, нормализация данных, LLM собирает аналитический отчёт по шаблону
- Команда получает готовый документ в нужном формате — вместо «собери сам из разных таблиц»
- 15 минут на полный отчёт — команда теперь анализирует данные, а не собирает их

Стек технологий

- OpenAI API
- Python · Jinja2
- REST API
- PostgreSQL
- Docker

Кейс: Аудит, который выявил потери за 2 года

[d.t]

Data Analytics • BI • Аренда и продажа автомобилей • Региональная сеть

100K+ записей
проанализировано

×10 быстрое принятие
решений

5 нед срок
выполнения

О проекте

- Данные о продажах, аренде и сервисе есть, но разрозненные — целостной картины нет
- Единый аналитический слой поверх трёх систем, автодашборды с ключевыми метриками
- Выявили узкое место в воронке аренды, которое незаметно съедало 15% выручки последние 2 года
- Управленческие решения на основе данных в реальном времени, не раз в квартал

Стек технологий

- Python · Pandas
- ClickHouse
- Apache Superset
- ETL pipeline
- REST API

Кейс: Карта освещённости города без ручной разметки [d.t]

ML • GIS • Спутниковые данные • Городская инфраструктура • Муниципальный заказчик

1200+ км² проанализировано

Авто обновление при
новых снимках

8 нед срок
выполнения

О проекте

- Нужно расставить приоритеты ремонта освещения на 1200 км² — вручную почти невозможно
- CV-модель классифицирует районы по спутниковым снимкам без участия людей
- Карта автоматически обновляется при поступлении новых снимков, приоритеты пересчитываются
- Обоснованный план ремонта вместо решений «на глаз», данные для защиты бюджета

Стек технологий

- Python · PyTorch
- GDAL · Rasterio
- PostGIS
- S3 Storage
- FastAPI

Кейс: Izzybooking — ИИ-поиск для тревел-агрегатора

[d.t]

AI • NLP • Рекомендации • Международный туристический маркетплейс • AI-агрегатор

100+ источников данных
в реальном времени

3x точнее, чем обычный
поиск

3 мес срок
выполнения

О проекте

- Пользователь вводит свободный запрос — система понимает намерение и собирает персональную подборку: отели, маршруты, гиды, активности
- NLP-модуль разбирает запрос на параметры поездки: даты, бюджет, тип отдыха, предпочтения — без сложных фильтров
- Рекомендательный движок ранжирует предложения от 100+ поставщиков в реальном времени по релевантности запросу
- Система обучается на поведении пользователей — с каждым запросом подборки становятся точнее

Стек технологий

- OpenAI API + LangChain
- Vector Search · Qdrant
- FastAPI · PostgreSQL
- NLP · Intent Parser
- Recommender Engine

Кейс: Ореол — ИИ-консьерж для навигации в медицине [d.t]

AI • Медицина • iOS/Android • Медицинский ассистент

RAG ответы строго по протоколам

AI оценка симптомов и маршрутизация

Проект в разработке с 2025 года

О проекте

- Помочь пациенту понять симптомы и решить это повод обратиться к врачу или нет
- Каждый ответ строится на клинических протоколах через RAG — никакой «творческой медицины» от модели
- Пользователь описывает симптомы → получает оценку ситуации → при необходимости записывается к врачу прямо в боте
- Единая команда закрыла весь стек: продукт, дизайн, backend, сайт и мобильное приложение



ОРЕОЛ

[ссылка на сайт](#)

[ссылка на приложение](#)

Стек технологий

- OpenAI API + RAG
- aiogram · Telegram
- React Native
- FastAPI · PostgreSQL
- Figma · UX/UI

Автоматизация обработки заявок и документов

[d.t]

NLP • Поиск • 1C

Сопоставление заявок с каталогом

389K позиций в каталоге

Оптовая дистрибьюция • Инженерная сантехника

- Заявки приходят в свободной форме: текст, фото, пересылки. 300 менеджеров искали вручную в каталоге на 389K позиций
- Гибридный поиск: по смыслу + по тексту + по структурным полям
- Оценка уверенности: высокая → корзина 1C, средняя → 2–3 варианта менеджеру, низкая → уточняющий вопрос

Извлечение заказов из документов

240 в бланке 1C

Производство под заказ • Противопожарные двери

- Заказы приходят фото, сканами, PDF и Excel — ручной перенос в 1C на сотни реквизитов
- Двухслойная система: извлечение значений с ссылкой на источник + движок бизнес-правил по регламенту
- Менеджер проверяет, а не перепечатывает: спорные значения в очередь, заказ ждёт подтверждения

Система машинного обучения для детектирования аномалий на авиационном производстве

[d.t]

ML • Мониторинг • IoT

-40%

времени на ручной
разбор показаний

0

пропущенных аномалий
за 8 месяцев

1000+

измерений за 24ч
в реальном времени

О проекте

- Тысячи измерений датчиков в сутки — аномалии замечали только после брака или поломки оборудования
- ML-модели, обученные на историческом браке — система сама пересчитывает пороги при изменении условий производства
- Инженеры получают алерт до поломки, а не после: переход от реактивного к превентивному обслуживанию
- — 40% времени на анализ, ноль критических пропусков за 8 месяцев промышленной эксплуатации

Стек технологий

- Python · scikit-learn
- Apache Kafka
- InfluxDB
- Grafana
- Docker · VPS

Три формата работы

[d.t]

Начинаем с малого: диагностика → пилот на ваших данных → полная разработка системы

Аудит и план

48–72 часа

Разбираем задачу, предлагаем архитектуру под неё и честно говорим, когда кастомная разработка оправдана, а когда нет.

- Анализ процесса и данных
- Архитектура решения под задачу
- Честно о рисках и ограничениях
- План пилота с критериями успеха
- Оценка сроков и бюджета

Пилот

2–6 недель

Разрабатываем рабочую систему на ваших данных. Вы видите метрики — потом решаете, масштабировать ли.

- 1–2 готовых рабочих сценария
- Тест на реальных данных и пользователях
- Интеграция в 1–2 ваших канала
- Передача сложных случаев оператору
- Отчёт с метриками и планом роста

Внедрение

4–12 недель

Полноценное кастомное ИИ-решение: разработка, тесты, интеграция, поддержка и развитие.

- Все сценарии и интеграции
- Доступы, аудит, безопасность
- Мониторинг качества в реальном времени
- Регламент обновления базы знаний
- Техническое сопровождение

Что мы используем в работе

[d.t]

Используем лучшие инструменты под вашу задачу — не привязываемся к одному вендору

Языковые модели

- OpenAI GPT-4o / o1
- Anthropic Claude
- YandexGPT · GigaChat
- Llama 3 (self-hosted)
- Fine-tuning · LoRA

RAG и поиск

- LangChain · LlamaIndex
- Qdrant · Weaviate · Chroma
- Elasticsearch (hybrid)
- pgvector
- Reranking · HyDE

Агенты и оркестрация

- LangGraph · CrewAI
- Tool use · Function calling
- Multi-agent системы
- Memory management
- Streaming

ML и данные

- Python · PyTorch
- scikit-learn · CatBoost
- scikit-learn · CatBoost
- Apache Kafka · Spark
- MLflow · DVC

Инфраструктура

- Docker · Kubernetes
- FastAPI · asyncio
- PostgreSQL · Redis
- AWS · YC · on-premise
- CI/CD · мониторинг

Интерфейсы

- Docker · Kubernetes
- FastAPI · asyncio
- PostgreSQL · Redis
- AWS · YC · on-premise
- CI/CD · мониторинг

Обсудим вашу задачу



dzeta.ai



contact@dzeta.ai



@Alexey_z_z



calendly/alex-dzeta



+7 929 818 5810

Ответим в течение 4 часов