



Balgabay Birzhan

Мужчина, 24 года, родился 16 мая 2001

+7 (747) 0519650 — предпочитаемый способ связи
birzhan981@gmail.com

Проживает: Алматы

Гражданство: Казахстан, есть разрешение на работу: Казахстан

Готов к переезду, готов к командировкам

Желаемая должность и зарплата

Python Backend/AI

2 000 \$ на руки

Специализации:

— Программист, разработчик

Тип занятости: полная занятость

Формат работы: гибрид, удалённо

Желательное время в пути до работы: не имеет значения

Опыт работы — 1 год 8 месяцев

Июнь 2025 —
Январь 2026
8 месяцев

TOO "LexARE"

Backend, AI Engineer

Разработка архитектуры RAG-систем:

- Спроектировал и внедрил архитектуру для семантического поиска по корпоративной базе знаний. Реализовал гибридный поиск (Vector + Full-text) для повышения точности релевантной выдачи.
- Разработал алгоритмы чанкинга (chunking) больших массивов данных для оптимизации контекстного окна LLM.

Мультиагентные системы и оркестрация:

- Построил масштабируемую систему автономных AI-агентов на базе LangChain, способную выполнять сложные цепочки задач (планирование, поиск, генерация ответа).
- Внедрил продвинутые техники промпт-инжиниринга (Chain of Thought, Few-shot) для улучшения качества рассуждений моделей.

Обработка данных (ETL и мультимодальность):

- Реализовал пайплайны для автоматического извлечения данных из неструктурированных документов (PDF, DOCX, XLSX) и изображений (OCR).
- Интегрировал модули распознавания голоса (ASR) для голосового управления и транскрипции встреч.

Октябрь 2024 —
Июнь 2025
9 месяцев

On. AI

AI разработчик

Инженер по AI-решениям / Специалист по интеграциям/ AI разработчик

Поддержка серверов, распределение клиентов по контейнерам и мониторинг нагрузки для обеспечения стабильной работы сервиса. Сбор требований и адаптация AI под специфику каждого бизнеса клиента. Разработка специфических интеграции и кастомный код под требования клиента, включая Prompt Engineering и составление документации.

Ключевые достижения:

Разработал и внедрил полнофункционального AI-ассистента с использованием LLM API, включая:

function calling

векторизацию данных

семантический поиск для повышения качества взаимодействия с клиентами.

Построил масштабируемую инфраструктуру с применением объектного хранилища S3, распознавания речи (SpeechKit), чтения PDF-документов и OCR (извлечение текста с изображений) для автоматизации документооборота.

Интегрировал AI с WhatsApp, Instagram и AmoCRM, обеспечив омниканальное взаимодействие с пользователями.

Реализовал передачу фото, видео и документов через AI, улучшив пользовательский опыт при работе с медиафайлами.

Настроил автоматическую интеграцию платежей по Kaspi QR с динамически поставляемой суммой.

Автоматизировал создание отчетов по результатам общения с клиентами, включая анализ и ключевые выводы.

Вёл прямую коммуникацию и переговоры с Enterprise-клиентами, B2B enterprise и SMB, оценивал стоимость и сроки проектов, предлагал 2–3 варианта реализации (ТЗ) с различным бюджетом и функциональностью для создания уникального торгового предложения (УТП)

Занимался продажей услуг, сопровождением и продлением контрактов за счёт качественного взаимодействия

Обучал стажёров и младших специалистов, организовывал внутреннее развитие команды.

Создавал видео тренинги для клиентов с обзором функционала и использованием платформы

Июнь 2024 —
Октябрь 2024
5 месяцев

TOO MyTrade(Outsourcing Company)

Python (Backend/ AI)

Разработка и сопровождение серверной части веб-приложений на микросервисной архитектуре.

Ключевые достижения:

Разработал мессенджера с нуля на стеке: FastAPI, WebSocket, SQLAlchemy + PostgreSQL, Docker, Kafka (Bitnami)

Обеспечил стабильную работу и масштабируемость сервиса.

Перевел функциональность монолитных компонентов в микросервисы, что снизило нагрузку и повысило отказоустойчивость. Использовал Docker, Kafka, FastAPI.

Интегрировал современные AI-инструменты для повышения интеллектуальности приложений:

OpenAI API (LLM)

Function calling, семантический поиск, векторизация данных

RAG (Retrieval-Augmented Generation)

Работа с S3-хранилищем, Speech Kit (speech-to-text)

Поддержка контекста диалога в интерактивных сценариях.

Образование

Высшее

2023
Высшее

Жетысуский государственный университет имени И. Жансугурова, Талдыкорган

Юриспруденция, Юриспруденция

Навыки

Знание языков	Казахский — Родной Английский — B2 — Средне-продвинутый Русский — C2 — В совершенстве
Навыки	<div><div>Docker</div><div>REST API</div><div>Английский язык</div><div>Python</div><div>FastAPI</div><div>Backend</div><div>LLM</div><div>Gitlab</div><div>gRPC</div><div>RAG</div><div>Векторизация</div><div>Анализ требований</div><div>Kubernetes</div><div>Оптимизация кода</div><div>Django Rest Framework</div><div>Django Framework</div><div>Pytest</div><div>Celery</div><div>RabbitMQ</div><div>CI/CD</div></div>

Опыт вождения

Права категории В

Дополнительная информация

Обо мне	Backend-разработчик + AI Engineer (ReAct Agents) Core: Python, FastAPI, SQLAlchemy, PostgreSQL. AI Stack: RAG, Prompt Engineering (CoT), ReAct Agents, Image/Audio/Files recognition (OCR/ASR). Infrastructure: Docker, Kubernetes, S3, Message Brokers (Kafka, RabbitMQ).
---------	---