

Әшекей Нұрмұхамед

Мужчина, 22 года, родился 30 апреля 2003

+7 (701) 1187021 — предпочитаемый способ связи • Github: <https://github.com/NurmukhamedKZ>

LinkedIn: <https://www.linkedin.com/in/nurmukhamed-ashekey-3031a3369/>

Telegram: @nureke3445

ashekeinureke@gmail.com

telegram: @nureke3445

Проживает: Алматы

Гражданство: Казахстан, есть разрешение на работу: Казахстан

Не готов к переезду, готов к командировкам

Желаемая должность и зарплата

AI-инженер

Специализации:

- Дата-сайентист
- Программист, разработчик

Тип занятости: полная занятость, частичная занятость, проектная работа/разовое задание, стажировка

Формат работы: на месте работодателя, гибрид, удалённо

Желательное время в пути до работы: не имеет значения

Опыт работы — 1 год 4 месяца

Сентябрь 2025 —
настоящее время
6 месяцев

Slideble (Slideshow generator)

AI Engineer

Сократил операционные расходы на генерацию слайд-шоу за счет оптимизации потребления токенов LLM на 30% через внедрение стратегий кэширования и продвинутого промпт-инжиниринга.

Ускорил процесс отладки и итерации промптов на 40% и значительно повысил надежность системы инициировав и интегрировав пайплайно bservability на базе LangSmith.

Исключил 100% ошибок рендеринга сцен вызванных некорректными ответами моделей спроектировав архитектуру пайплайна структурированного вывода данных с жестким соблюдением JSON схем.

Повысил фактологическую точность генерируемого контента проведя сравнительный анализ поисковых инструментов и интегрировав Tavily API для извлечения внешнего контекста.

Tech Stack: Python, OpenAI API, LangChain, LangGraph, LangSmith, PostgreSQL, SQLAlchemy, Tavily API. FastAPI

Ноябрь 2024 —
Сентябрь 2025
11 месяцев

JauapAI (Educational RAG Platform)

AI Engineer

Увеличил показатель удержания пользователей снизив время получения первого токена TTFT на 62% с 40 до 15 секунд через внедрение асинхронной обработки и тюнинг векторного поиска.

Решил проблему точности поиска сложной академической терминологии внедрив Hybrid Search на базе Dense и Sparse эмбеддингов что позволило устранить критические

галлюцинации модели.

Автоматизировал оцифровку сложных учебных материалов с формулами и таблицами разработав ETL пайплайн с использованием VLM воркфлоу и стратегий Semantic Chunking.

Обеспечил высокую доступность системы и поддержку конкурентных сессий спроектировав масштабируемый бэкенд на FastAPI и Docker

Tech Stack: Python, FastAPI, Docker, LangChain, Qdrant (VectorDB), Redis, PostgreSQL, OpenAI API, Telegram API. ML Ops & Embeddings: Hybrid Search (BGE-M3, Gemini-001, Voyage-4).

Образование

Бакалавр

2025
Бакалавр

Казахстанско-Британский технический университет, Алматы
Школа Информационных Технологий и Инженерии, Информационные системы

Навыки

Знание языков
Казахский — Родной
Английский — C1 — Продвинутый
Русский — C2 — В совершенстве

Навыки
Python SQL Английский язык PostgreSQL Git MySQL langchain
NLP LLM PyTorch pandas Scikit-learn Docker FastAPI RAG
ML MLflow AI agent Matplotlib Seaborn AI TensorFlow API
Deep Learning AI engineer ChatGPT PROMT Numpy

Дополнительная информация

Обо мне
Я AI инженер с фокусом на разработку масштабируемых RAG систем и автономных AI агентов которые решают реальные задачи бизнеса.

Telegram: @Nureke3445