

# Diyar Makhambetali

🌐 [Portfolio](#) | ✉ [Email](#) | [LinkedIn](#) | [GitHub](#)

## EDUCATION

---

### Nazarbayev University

*Bachelor of Science in Computer Science*

August 2024 - June 2028 (Expected)

*Astana, Kazakhstan*

## EXPERIENCE

---

### Institute of Smart Systems and Artificial Intelligence (ISSAI)

*Backend Engineer*

December 2025 – Present

*Astana, Kazakhstan*

- Architected and shipped a standalone **audio separation** (speech/noise) service for **ISSAI Voice Studio**, a Kazakh **TTS** & voice-cloning platform, powered by Meta's **SAM Audio** model, served via **FastAPI**, containerized with **Docker**, and persisting output to **Garage S3**.
- Built a standalone **audio watermarking** microservice based on **SilentCipher (FastAPI, Docker)**, embedding identifiers into generated clips to enforce commercial-use limits; load-tested with **Locust** at up to **~135 req/s** and **~20 ms** latency.
- Secured the platform's microservices with **Keycloak**, implementing centralized authentication and role-based access control.
- Engineered an async **voice generation** pipeline (**TaskIQ + RabbitMQ**) chaining an **LLM** enrichment step with a Kazakh **TTS** service, decoupling long-running generation from API requests.
- Migrated the production **LLM** serving stack from **vLLM** to **SGLang** with tensor parallelism across multiple **GPUs**, cutting **VRAM** consumption by **68%** and freeing capacity to co-host other services and lower inference cost.
- Diagnosed and resolved throughput bottlenecks in the platform's async **Django + RabbitMQ + Celery** architecture via horizontal scaling of **Docker**-based consumers, reducing queue backlog and improving **LLM** streaming response latency.
- Designed and implemented a **CI/CD pipeline** with automated **Pytest** execution on **self-hosted** infrastructure, reducing manual testing effort by **70%**.

### Baiterek National Investment Holding

*Intern Data Analyst*

September 2025 – October 2025

*Astana, Kazakhstan*

- Automated extraction of **KASE** bonds, trades, and coupons into **CSV** using **Python**, **Playwright**, **pandas**, and **n8n** workflows.
- Built a fully automated **data pipeline** with **async programming**, reducing scraping time from **1 hour to under 10 minutes**.
- Built **BI** dashboards in **Tableau** and **Metabase**, developing **ETL** pipelines with complex **raw SQL** queries for analytics and financial reporting.

### Clario AI

*Python Backend Developer (Full-time)*

April 2025 – August 2025

*Astana, Kazakhstan*

- Integrated **STT** services (**ElevenLabs** and **AssemblyAI**) for automated processing of phone call recordings.
- Integrated **CRM** systems (**AmoCRM** and **Bitrix24**) via **webhooks** for real-time lead status tracking, saving an estimated **15+ hours** of manual work per week.
- Architected a **Gemini LLM**-based deal revival system that analyzed failed leads from **PostgreSQL**, driving re-engagement of **200+ leads** and a **15% increase** in conversions from that segment.
- Implemented **JWT** authentication and **OAuth 2.0** to secure API endpoints.
- Integrated **Redis** caching for frequently accessed lead data, reducing average API response time by **45%**.

## TECHNICAL SKILLS

---

**Languages:** Python, TypeScript, SQL

**Backend:** Django, DRF, FastAPI, Celery, TaskIQ

**Databases:** PostgreSQL, MySQL, Redis

**DevOps:** Docker, Nginx, CI/CD, Pytest, Locust

**AI/ML:** LLM, vLLM, SGLang, faster-whisper, STT/TTS

**Other:** Garage S3, RabbitMQ, Keycloak, Git

## AWARDS

---

**KazMunaiGaz Hackathon:** 2nd / 54 (2026)

**Outpeer Data Science:** 2nd / 80 (2025)

**Decentrathon 4.0:** Top 10 / 300 (2025)  
*National hackathon*