

AZIZKHUJA ISLOMKHUJAEV

Backend AI Engineer

azizkhuja_isl@outlook.com | +998 93 566 24 30 | Tashkent, Uzbekistan

linkedin.com/in/azizkhuja-islomkhujaev | github.com/Azizhoja

SUMMARY

Backend-focused AI Engineer with **1.5+ years of professional experience** building LLM-powered applications, RAG pipelines, tool-calling agents, and production backend services. Experienced in LoRA fine-tuning, document/OCR workflows, TTS integration, cloud/local LLM deployment, and FastAPI-based AI systems. Strong **Python** and **Java** foundation, with hands-on work across AI tutoring platforms, media intelligence pipelines, backend automation, and database-backed applications. Ranked 1st in Software Engineering cohort at IT Park University.

SKILLS

AI & LLM Systems: LLM Applications, RAG Pipelines, Tool-Calling Agents, LoRA Fine-Tuning, Hugging Face Transformers, PyTorch, LangChain, CLIP, SmolDocling, TTS Integration, Local LLM Deployment

Backend Development: Python, Java, FastAPI, Flask, Quarkus, REST APIs, PostgreSQL, MongoDB

Infrastructure & Tools: Linux, Docker, Git, GitLab CI/CD, Gunicorn/Uvicorn, Azure

Practices: Backend Architecture, API Design, Database Design, CI/CD Automation, Agile/Scrum

PROFESSIONAL EXPERIENCE

Salen

12/2024 – 04/2026

AI Engineer & Backend Developer

Tashkent,
Uzbekistan

- Developed backend services for Salen Academy, an AI-powered math tutoring platform combining LLM based reasoning, Wolfram Alpha tool integration, Manim-generated explanatory videos with multilingual TTS output.
- Built and maintained FastAPI services for AI workflows, including RAG, tool-calling agents, and sandboxed code execution.
- Worked on LoRA-based LLM fine-tuning for math reasoning and explanation tasks, improving the quality of generated responses.
- Improved document/OCR processing accuracy by 41% on internal evaluation data through model selection, and hyperparameter tuning.
- Reduced average inference cost per request by 40% by implementing on-demand GPU workers and cloud deployment optimization across Azure.
- Improved CI/CD deployment speed by 19% through Docker-based service packaging, Uvicorn worker configuration

PROJECTS

Video Intelligence App

Personal Project

- Designed and built an automated media intelligence pipeline for processing short-form video content from public or user-provided sources.
- Implemented video ingestion, frame/audio extraction, semantic classification using OpenAI CLIP, and speech transcription using NVIDIA Parakeet-TDT-0.6B-v3.
- Built searchable metadata outputs for content discovery, semantic filtering, and media library organization.
- Focused on practical backend pipeline design, batch processing, model integration, and AI-driven media analysis.

Equipment & User Management System

University Course Project

- Developed a Java-based command-line equipment management system with user management, equipment tracking, Layered Architecture, CRUD operations, and PostgreSQL-backed persistence.
- Designed database tables, user workflows, and CLI interactions for managing equipment records and user access.

EDUCATION

BSc. Software Engineering

2024 – 2028

IT Park University / EPAM Campus

GPA: 4.9 / 5.0 · Full Scholarship

Focus: Backend Java Development, Databases, Operating Systems, Information Security.

COURSES

Hugging Face Transformers Course

Hugging Face

- Gained practical experience in Transformer models, fine-tuning LLMs, and building NLP applications using the Transformers, Datasets, and Accelerate libraries.

Python 3 Deep Dive

Udemy - Dr. Fred Baptiste

- Learned Python internals, memory management, functional programming, iterators, generators, and advanced object-oriented programming.

Python API Development

freeCodeCamp.org

- Gained hands on experience in production-ready REST API using FastAPI, including Pydantic models, PostgreSQL with SQLAlchemy, JWT authentication, Alembic migrations, Docker, testing with pytest.

LANGUAGES

: Uzbek - C2, Russian - C2, English - C1