Academic Curriculum Vitae

Khasankhon Yusupkhujaev
Data Scientist | Data Engineer
Tashkent, Uzbekistan
Email: **khasanya99@gmail.com**
Phone: +998900131115
LinkedIn: linkedin.com/in/hasan-yusufkhuja-22a516162
Personal website: https://khasanyusupkhujaev.github.io/
GitHub: https://github.com/khasanyusupkhujaev?tab=repositories

---

Research Interests

Natural Language Processing (NLP), Machine Learning (ML), Reinforcement Learning, Low-resource Language Models, Human-centered AI, Social Impact of AI

Research Outputs (in preparation)

Benchmarking Pre-Trained Open-Source Large Language Models for Uzbek: Evaluating Performance in a Low-Resource Setting Across Translation, Comprehension, and Generation (manuscript in preparation).

Predicting Match Outcomes in The Hundred Cricket Format: A Machine Learning Approach? MSc Dissertation, Durham University (2024).

Education

Durham University — *MSc in Data Science* (2023–2024)

- Awarded **Distinction**.
- Dissertation: *Can We Predict the Outcomes of Cricket Matches?*
- Fully funded by the **El-Yurt Umidi Foundation Scholarship**, a competitive government program covering all tuition and living expenses for postgraduate study abroad.

INHA University in Tashkent — *BBA in Logistics* (2018–2022)

Research Experience

MSc Dissertation — Durham University (2024) Paper Code

- Built predictive models for cricket match outcomes using advanced ML techniques.
- Engineered features such as win streaks, venue effects, and head-to-head stats.
- Explored fairness and politics of data production in sports analytics.

Uzbek LLM Benchmarking Project (2025) Paper Code

- Designing the first **comprehensive benchmark for Uzbek large language models**.
- Developing evaluation protocols (BLEU, LLM evaluation, human evaluation) for low-resource NLP.
- Preparing open-source framework for reproducible benchmarking.

Uzbek LLM (from Scratch) (2023–2024)

- Collected and cleaned a large-scale Uzbek corpus.
- Trained transformer-based architectures tailored for low-resource settings.
- Conducted systematic evaluation against multilingual baselines.

Smart Cradle (Beshique) Project (2024–2025) Code

- Developed Raspberry Pi–based cradle detecting infant wakefulness via computer vision and audio analysis.
- Applied reinforcement learning for adaptive soothing mechanisms.

Academic Projects

Teacher LLM — Domain-specific model assisting educators with content creation and curriculum support. Code

Supreme Court Decision Predictor — Applied NLP to predict judicial outcomes based on precedent texts. Code

Treasury Semantic Search — Built semantic retrieval system for government documents. Code

---

Scholarships & Awards

El-Yurt Umidi Foundation Scholarship (2023–2024) — Government of Uzbekistan program funding study abroad (BSc, MSc, PhD, internships). Fully covered MSc studies at Durham University.

Professional Experience *(condensed)*

Data Scientist & Data Engineer — Ministry of Digital Technologies (2025–Present)
Lead Specialist in Data Analytics — Digital Government Center (2024–2025)
SAP ABAP Developer & Master Data Manager — Korzinka (2021–2023)

Certifications

- IBM — *Develop Generative AI Applications: Get Started*
- IBM — *Machine Learning with Python*
- IBM — *Data Science Professional Certificate*
- DeepLearning.AI — *Machine Learning*