

Abror Shopulatov

✉ murodbeck@proton.me  [murodbecks.github.io](https://github.com/murodbecks)  github.com/murodbecks  huggingface.co/murodbek

AI Engineering student at MBZUAI with hands-on experience in NLP and machine translation. Contributed to Uzbekistan's largest open-source text corpus and developed production ML pipelines at Tahrirchi. Strong foundation in mathematics with a passion for advancing language technology for low-resource languages.

Job Experience

Machine Learning Engineer (<i>Tahrirchi</i>)	Tashkent, Uzbekistan 02/2023 - present
<ul style="list-style-type: none">Involved in the collection and preprocessing of Uzbekistan's largest text corpus, enhancing skills in large-scale data handling and facilitating its open-source release.Designed and implemented an end-to-end ML pipeline for machine translation, gaining experience in parallel data collection and advanced cleaning techniques.Managed synthetic data generation workflows and developed a context-based translation feature, improving model performance and usability similar to Reverso.	

Education

Bachelor of Science in Artificial Intelligence <i>MBZUAI</i>	Abu Dhabi, UAE 2025-2029
--	------------------------------------

AI Engineering major at Mohamed bin Zayed University of Artificial Intelligence, a graduate-level research university specializing in AI.

Publications

- Shopulatov, A.** (2025). *UzLiB: A Benchmark for Evaluating LLMs on Uzbek Linguistics*. Tilmoch.ai. (Blog Post). tilmoch.ai (in Uzbek)
- Mamasaidov, M., & **Shopulatov, A.** (2024). Open Language Data Initiative: Advancing Low-Resource Machine Translation for Karakalpak. In *Proceedings of the Ninth Conference on Machine Translation (WMT24)* (pp. 606-613). Association for Computational Linguistics. (Conference Paper). DOI: 10.18653/v1/2024.wmt-1.48 / [Link](#)
- Mamasaidov, M., **Shopulatov, A.**, & Yusupov, J. (2023). *How we created an Uzbek grammar checker*. Tahrirchi. (Blog Post detailing dataset/model release). tahrirchi.uz (in Uzbek)

Skills

- Programming:** Python
- AI/ML Libraries & Frameworks:** PyTorch, HuggingFace Ecosystem (transformers, datasets, tokenizers), Scikit-learn, MosaicML
- AI/ML Concepts:** Natural Language Processing (NLP), BERT, Seq2Seq Models, Machine Translation, Synthetic Data Generation, Transfer Learning, Data Preprocessing, Model Training & Evaluation, Text Corpus Development
- Data Science Tools:** Pandas, NumPy, Matplotlib, Data Collection & Cleaning
- Platforms & Tools:** Google VertexAI GenAI SDK, Git
- Soft Skills:** Problem-Solving, Research, Open-Source Contribution, Project Initiative, Cross-functional Collaboration

Languages

- Uzbek [Native]
- English [Advanced]