Jafar Isbarov

Blacksburg, VA isbarov@vt.edu ceferisbarov.github.io ceferisbarov Google Scholar

Summary

1st year PhD student working on trustworthy AI. My current research focus is reliability of LLM outputs. Previously, I worked on low-resource language modeling and evaluation. I have 3 years of industry experience building production NLP systems, including chatbots and spelling correction software.

Education

Virginia Tech *Aug. 2025 –*
PhD in Computer Science

- Conducting research in The Data Security and Privacy Lab under supervision of **Murat Kantarcioglu**.
- Collaborating with **dlab@EPFL** on structured generation and grammar-aligned decoding.

The George Washington University *Aug. 2023 – May 2025*
MS in Computer Science

- **GPA:** 3.92
- **Thesis:** Evaluating Language Understanding and World Knowledge of Large Languages Models in Turkic Languages

ASOIU *Aug. 2018 – May 2023*
BS in Biomedical Engineering

- **Thesis:** Computational Methods for the Analysis of Single-Cell Transcriptomics Data

Experience

Machine Learning Engineer *Baku, Azerbaijan*
eiLink R&D *Jan. 2025 – Aug. 2025*

- Developed an agentic chatbot for querying drilling datasets, enabling natural language interaction with heterogeneous log formats for **British Petroleum**.
- The chatbot reduced typical data analysis and reporting tasks from 2-8 hours to 15-30 minutes, delivering 10-20x efficiency improvements.

Research Affiliate *New York, NY*
New York University *July 2024 – Dec. 2024*

- Worked on an adversarial paradigm for prompt compression with Duygu Ataman.
- Built TUMLU - the first native multilingual language understanding benchmark for Turkic languages. [1]

Machine Learning Engineer *Baku, Azerbaijan*
Prodata LLC *Mar. 2023 – Jul. 2024*

- Developed a new technique for hierarchical document retrieval in the legal domain. [2]
- Developed monolingual encoder-only foundation models which outperformed the multilingual SOTA models for Azerbaijani language understanding by 0.5%. [3]

Machine Learning Engineer *Baku, Azerbaijan*
Azerbaijan Artificial Intelligence Lab *Jul. 2022 – Mar. 2023*

- Developed a robust spelling correction method for agglutinative languages, reducing Word Error Rate from 27.00% to 17.67%. [4]

Research Intern *remote*
Helmholtz Zentrum München *Jan. 2021 – May 2021*

- Conducted a comparative analysis of dimensionality reduction and clustering algorithms for single-cell transcriptomic data analysis (with **Elmir Mahammadov**).

Publications

[1] TUMLU: A Unified and Native Language Understanding Benchmark for Turkic Languages Aug. 2025

Jafar Isbarov, Arofat Akhundjanova, Mammad Hajili, et al.

ACL 2025, Main Conference

[2] Enhanced document retrieval with topic embeddings Sep. 2024

Jafar Isbarov, Kavsar Huseynova

IEEE AICT 2024

[3] Open foundation models for Azerbaijani language Aug. 2024

Jafar Isbarov, Kavsar Huseynova, Elvin Mammadov, et al.

ACL 2024, 1st SIGTURK Workshop **(Honorable Mention in the Best Paper Category)**

[4] Robust Automated Spelling Correction with Deep Ensembles Aug. 2024

Jafar Isbarov, Kavsar Huseynova, Samir Rustamov

ACM ISMSI 2024

Skills

Programming: Proficient with Python, comfortable with Julia and C++.

ML/AI: Traditional ML (scikit-learn). Low-level and high-level deep learning libraries (PyTorch, TensorFlow, Transformers). Pre-training and efficient fine-tuning of LLMs. Specifically worked with BERT, GPT, and LLaMA architectures.

Professional Service

Reviewer at:

- ACL ARR 2025 February
- 1st LLMSEC Workshop (ACL 2025)
- 4th Multilingual Representation Learning Workshop (EMNLP 2024)
- Neural Computing and Applications Journal (2024)

Officer at ACL Special Interest Group in Turkic Langauges (SIGTURK)

Extracurricular Activities

Open-source contributions to:

- Hugging Face Ecosystem (lighteval, Python Client)
- Julia Ecosystem (TableTransforms.jl, Julia-LLM-Leaderboard, CounterfactualExplanations.jl)