Jafar Isbarov

jafar.isbarov@gwu.edu   ceferisbarov.github.io   U.S. (F1 visa)

Skills

C++, CUDA, Python, Julia, PyTorch, TensorFlow, FastAPI, Rasa, PostgreSQL, Docker, Linux, GCP, System Design

Education

George Washington University, M.Sc. Computer Science, Expected: 2025, CGPA: 3.85 (1st of class)
ASOIU, B.Sc. Biomedical Engineering, 2023. CGPA: 3.4

Experience

Research Affiliate with Duygu Ataman @ New York University - New York City, NY   Aug. 2024 –
:   System-level evaluation of existing prompt compression methods.
:   Developing an adversarial prompt compression paradigm.
:   Building TurkicMMLU - the first native multilingual language understanding benchmark for Turkic languages.

Lead machine learning engineer @ PRODATA LLC - Baku, Azerbaijan   Mar. 2023 – Jul. 2024
:   Built and led the company's NLP team and ML infrastructure.
:   Led the development of chatbot systems in Azerbaijani.
:   Designed and led the development of an internal on-premise no-code solution for chatbot development. (Stack: Rasa, Django, Nuxt, Celery, PostgreSQL, and MinIO)
:   Developed an improved document retrieval strategy for hierarchical text corpora.
:   Developed the SOTA monolingual foundation models for Azerbaijani.

Machine learning engineer @ Space Agency of the Republic of Azerbaijan   Jul. 2022 – Mar. 2023
:   Built a new robust spelling correction algorithm for agglutinative languages, decreasing Word Error Rate from 27.00 to 17.67.
:   Built the backend system for the spelling correction solution (Stack: MariaDB, FastAPI, TensorFlow).
:   Managed the data annotation team for the spelling correction project.
:   Built a prototype chatbot to replace the largest customer service system (BirBank) in Azerbaijan.

Teaching Assistant @ QSS Analytics - Baku, Azerbaijan   May 2021 – Nov. 2021
:   TA in Introduction to Python and Big Data classes.
:   Grading homework, and out-of-class tutoring for students.
:   Gave extra lectures for underperforming students.

Research Assistant with Elmir Mahammadov @ Helmholtz Zentrum München - Remote   Jan. 2021 – May 2021
:   Performed literature review on existing methods for single-cell transcriptomic data analysis.
:   Reproduced two papers and performed a comparative analysis of dimensionality reduction, clustering, and visualization algorithms.
:   Wrote a detailed report summarizing our findings and making recommendations for each stage of the single-cell transcriptomic data analysis pipeline.

Research output

- : **J. Isbarov**, K. Huseynova, E. Mammadov, M. Hajili, D. Ataman, "Open foundation models for Azerbaijani language", *ACL 2024, 1st SIGTURK Workshop,* **Honorable Mention in Best Paper category,**
https://aclanthology.org/2024.sigturk-1.2/
- : **J. Isbarov**, K.Huseynova, "Enhanced document retrieval with topic embeddings," *IEEE AICT 2024,*
https://ieeexplore.ieee.org/document/10740455
- : **J. Isbarov**, K. Huseynova, S. Rustamov, "Robust Automated Spelling Correction with Deep Ensembles", *ACM ISMSI 2024,*
https://dl.acm.org/doi/10.1145/3665065.3665070
- : D. Ataman, ..., **J.Isbarov** et al. "Findings of the 2nd Shared Task on Multi-lingual Multi-task Information Retrieval at MRL 2024," *[Ongoing],*
https://sigtyp.github.io/ws2024-mrl.html
- : **J. Isbarov**, E. Mahammadov, "Comparison of algorithms used in single-cell transcriptomic data analysis", *Preprint,* 2021, https://arxiv.org/abs/2408.12031

Reviewer activity

Reviewer @ 4th Multilingual Representation Learning Workshop (EMNLP 2024)
Reviewer @ Neural Computing and Applications Journal

Miscellaneous roles and achievements

Researcher @ ACL Special Interest Group on Turkic Languages
Organizer @ 4th Multilingual Representation Learning Workshop (EMNLP 2024)
Presenter @ DevFest 2023 Baku: Machine Learning at Scale
Presenter @ GDG Baku: The Challenges of the Azerbaijani Language in Artificial Intelligence
Co-founder @ aLLMA Lab, where we develop open-source NLU technologies for low-resource languages
Open-source contributor @ TableTransforms.jl, HuggingFace Python Client, etc.