Mohammad (Hossein) Moslemi

Trustworthy Machine Learning Researcher

mhmoslemi2338@gmail.com • Homepage • Canada (Open to relocate)
GitHub • LinkedIn • Google Scholar

Professional Summary

Responsible AI researcher with an M.Sc. in Computer Science and a B.Sc. in Engineering. I focus on **fairness in AI** and have expertise in **causality**. Skilled in Python, I build research pipelines, fine-tune LLMs, and have hands-on experience as a **data engineer**. Actively contribute to open-source (1.1k GitHub stars, recipient of a **Microsoft award**). I have presented and published at top venues, including **SIGMOD and IEEE BigData**. Currently seeking opportunities in **applied machine learning** and **trustworthy AI**.

Technical Core Skills

- **Programming:** Python (advanced, open source contributor), C++, MATLAB

- **Frameworks:** PyTorch, TensorFlow, NumPy, Pandas, scikit-learn, SciPy, POT (Optimal Transp.)

- **Machine Learning:** Deep Learning, Gradient Boosting, Fairness, Optimal Transport, Causality

- **NLP:** LLM fine-tuning (Wav2Vec2, Persian language models), Transformers

- **Data Engineering:** PySpark, PostgreSQL, Redis, SQL

- **MLOps & Tools:** Docker, Git, GitHub Actions, LaTeX, Bash, Linux

- **Communication:** Technical writing (publications), conference presentations, and teaching

Selected Publications & Research Highlights

- Reducing Bias in Record Matching via Score Calibration, **IEEE Trans.**, 2025 (in review).

- *Lead author (1/3)*, Evaluating Blocking Biases in Entity Matching, **IEEE BigData**, 2024.

- *Second author (2/5)*, OTClean: Data Cleaning Using Optimal Transport, **SIGMOD**, 2024.

- Threshold-Independent Fair Matching through Score Calibration, **SIGMOD Workshop**, 2024.

- Contributor to Samila (open-source generative art, 1.1k GitHub stars), **Microsoft award**.

- Presented technical talk and poster at **SIGMOD** to academic and industry audiences.

Education

2023 – Apr. 2025	**MSc in Computer Science**, Western University, London, ON, Canada *Thesis:* Fairness in Entity Matching and Blocking *Supervisor:* Dr. Mostafa Milani
2018 – 2023	**BSc in Electrical Engineering**, Sharif University of Tech., Iran

Professional Experience

Fair Classification via Score Calibration and Optimal Transport   Sep. 2023 – Mar. 2025

- Initiated and led a **fair record linkage** project, developing a bias removal method by aligning the matching score distributions of demographic groups using optimal transport.
- Defined new fairness metrics for clustering and demonstrated bias in benchmarks and downstream tasks, presented results at **IEEE BigData'24**.
- Built an **optimization framework** to reduce demographic parity and label-dependent biases in LLM-based record linkage, and evaluated the impact of noise, missing data, synonyms (via BERT), and hierarchies, presenting at **SIGMOD'24**.

Conditional Independence in Data via Optimal Transport   Sep. 2023 – Feb. 2024

- Applied optimal transport to clean datasets violating causal relationships, enabling **bias removal**.
- Collaborated with teams in Canada and top U.S. universities, leading essential implementation and experiments for a **SIGMOD'24** paper, and presented the work at the conference.

Data Engineering Intern – Streaming & Distributed Systems   Sep. 2022 – Jan. 2023

- Initially developed and maintained backend services using **FastAPI**.
- Recognized for strong work ethic and potential and was moved to the **data engineering team**.
- Built scalable **ETL pipelines** with PySpark for real-time cleaning of clickstream and behavioral data, and automated workflows to ensure reliable, efficient data processing.
- Assisted **DevOps teams** in migrating architecture to Kubernetes for reliability and efficiency.

Persian Speech-to-Text Messenger Bot (Wav2Vec2)   Jan. 2022 – Sep. 2022

- Fine-tuned `wav2vec2` on ~350 h of in-house Persian audio at an AI startup.
- Built and deployed a Telegram Messenger bot that transcribes user voices to Persian text via ASR API (≤ 2 s median latency) and integrated inline product catalog and payment links.