Mohammad (Hossein) Moslemi

Trustworthy Machine Learning Researcher

mhmoslemi2338@gmail.com • 226-998-5260 • Canada (Open to relocate)
Homepage • GitHub • LinkedIn • Google Scholar

Professional Summary

Responsible AI researcher with an M.Sc. in Computer Science and a B.Sc. in Engineering. Focused on **fairness in AI**, with expertise in causality. Skilled Python developer with experience building research pipelines, fine-tuning LLMs, and working as a **data engineer**. Active in the open-source community (1.1k GitHub stars, **Microsoft award**). Presented and published in top venues (**SIGMOD & IEEE BigData**). Seeking opportunities in **applied ML** and **trustworthy AI**.

Technical Core Skills

- **Programming:** Python (advanced, open source contributor), C++, MATLAB
- **Frameworks:** PyTorch, TensorFlow, NumPy, Pandas, scikit-learn, SciPy, POT (Optimal Trans.)
- **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 Match. through Score Calib., under review **IEEE Transactions**, 2025.
- *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-Indep. Fair Match. 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. M. 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 matching score distributions using optimal transport.
- Defined new fairness metrics for the clustering step and proved bias in benchmarks and downstream tasks; presented results at **IEEE BigData 2024**.
- Built an **optimization framework** to reduce demographic parity and label-dependent biases in LLM-based record linkage, evaluating the impact of noise, missing data, synonyms (via BERT), and hierarchies; presented at **SIGMOD 2024**.

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

- Applied optimal transport to repair datasets violating causal relationships for **bias removal**.
- Collaborated with teams in Canada and top U.S. universities, leading implementation and experiments essential for **SIGMOD'24** acceptance and presenting 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, moved to the **data engineering team**.
- Built scalable **ETL pipelines** with PySpark for real-time cleaning of clickstream and behavioral data. 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 ($\leq$2 s median latency); integrated inline product catalog and payment links.