Mohammad (Hossein) Moslemi

226-998-5260 • mohammad.moslemi@uwo.ca • Google Scholar

Research Interests: Responsible and Trustworthy AI. I study algorithmic bias in foundation models, using optimal transport and distributional alignment techniques to remove bias in LLM outputs. I'm especially interested in adapting deep learning pipelines to respect fairness constraints.

Education

Sep. 2023 – Apr. 2025 **MSc in Computer Science**
Western University, London, ON, Canada
Thesis: *Fairness in Entity Matching and Blocking*
Supervisor: Dr. Mostafa Milani

Sep. 2018 – Sep. 2023 **BSc in Electrical Engineering**
Sharif University of Technology, Tehran, Iran
Thesis: *Lung Tissue Classification via Graph Signal Processing on CT-Scans*
Supervisor: Dr. Arash Amini

Publications

Conference

C1. **Moslemi, M. H.**, Balamurugan, H. & Milani, M. Evaluating Blocking Biases in Entity Matching. *IEEE International Conference on Big Data **(IEEE Big Data)*** (2024).

C2. Pirhadi, A., **Moslemi, M. H.**, Cloninger, A., Milani, M. & Salimi, B. Otclean: Data cleaning for conditional independence violations using optimal transport. *Proceedings of the ACM on Management of Data **(SIGMOD)*** **2,** 1–26 (2024).

Workshop

W1. **Moslemi, M. H.** & Milani, M. *Threshold-Independent Fair Matching through Score Calibration* in *Proceedings of the Conference on Governance, Understanding and Integration of Data for Effective and Responsible AI Workshop at **SIGMOD*** (2024), 40–44.

Under Review

S1. **Moslemi, M. H.** & Milani, M. Reducing Biases in Record Matching through Scores Calibration. *IEEE Transactions on Knowledge and Data Engineering* (2025).

S2. **Moslemi, M. H.**, Mousavi, A., Behkamal, B. & Milani, M. Entity Matching and Data Heterogeneity: Survey and Experimental Analysis. *Journal of Data and Information Quality* (2025).

S3. **Moslemi, M. H.**, Omati, M. M. & Amini, A. Lung Tissue Classification for ILD patients using Graph Signal Processing. *IEEE Transactions on Biomedical Engineering* (2025).

Presentations

T1. **Moslemi, M. H.** & Milani, M. *Threshold-Independent Fair Matching through Score Calibration* Governance, Understanding and Integration of Data for Effective and Responsible AI Workshop at **SIGMOD** (Santiago, Chile). June 2024.

Professional Experience

Fair Matching via Score Calibration Sep. 2024 – Mar. 2025

- Led the project building on earlier SIGMOD work, fully formalized the fair classificaion problem, and used a faster dual approach to optimal transport that gives the same results.
- Designed two algorithms to reduce both demographic parity and label-dependent biases such as equalized odds.

Institutional Affiliation and Citation Bias Nov. 2024 – Apr. 2025

- Collected a large citation dataset with web-scraping and cleaned it automatically in PySpark.
- Showed how author affiliation skews citation counts across research fields.

Record Linkage under Heterogeneous Data Jul. 2024 – Dec. 2024

- Studied how noise, missing data, synonyms (injected using BERT), and hierarchies hurt record-linkage benchmarks.
- Ran controlled experiments on several datasets to measure train-test performance drops.

Blocking Bias Analysis in Record Linkage Apr. 2024 – Aug. 2024

- Initiated and led the project by drawing a novel connection between blocking in record linkage and clustering, focusing on fairness implications.
- Defined new fairness metrics and proved bias in standard benchmarks and downstream tasks.

Bias Removal via Optimal Transport Sep. 2023 – Apr. 2024

- Initiated and led the project, formulating a novel bias removal method by aligning the classification score distributions of majority and minority groups via optimal transport.
- Developed the optimization framework and implemented it. Applied it to LLM-based record linkage using post-processing. Presented the resulting paper at the SIGMOD 2024.

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

- Used optimal transport (Sinkhorn) to fix datasets that violate causal conditional independence.
- Joined during the later stages of the project; led the implementation and experiments that were crucial for the acceptance of the paper at the SIGMOD 2024.

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

- Built scalable ETL pipelines using `PySpark` and `Kafka` to process streaming data in real-time.
- Deployed containerized services via `Kubernetes`, improving system reliability and deployment efficiency across clusters.

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); integrated inline product catalog and payment links.

Key Skills

Technical

- **Python (advanced):** POT (Optimal Transport Lib), NumPy, SciPy, Pandas, scikit-learn, TensorFlow, PyTorch
- Git & GitHub • Linux / Bash

Soft Skills

- **Project management:** Run multiple research projects and hit deadlines.
- **Scientific communication:** Publish papers, review manuscripts, present at conferences.
- **Self-Discipline:** Strong self-motivation and discipline, consistently meeting deadlines and goals.
- **Resilience & Perseverance:** Never quits under pressure, consistently completes tasks regardless of challenges.

Academic Service

Sep. 2023 – Mar. 2025 **Teaching Assistant**, Computer Science department, Western University

- Python Teaching Assistant for three semesters, teaching additional concepts and ensuring student comprehension.
- Managed and graded lab sessions for 50 students.

Jun. – Sep. 2024 **Academic Advisor**, Computer Science Dep., University of Western Ontario

- Guided and assisted Master of Engineering students with their directed studies and projects.

Relevant Coursework

Machine Learning Advanced Deep Learning • Statistical learning • Reinforcement Learning • AI & Bio-Computing

Data Management Advanced Databases • Speech Recognition • Data Structures & Algorithms • Big Data Analysis

Mathematics Linear Algebra • Digital Signal processing • Probability and Statistics

Honors and Certificates

2025 Master's Thesis Commended as a **New Standard** by Defense Committee

2023 Ranked in the top 10% among the class of 2023

2023 Astonishing Achievement Award for the B.Sc. Project

2017 **Bronze medal** in National Physics Olympiad