

# Amirreza Naziri

 [Amir79Naziri](#) |  [amirreza-naziri](#) |  [website](#) |  [+1\(437\)818-8378](#)  
 [naziriamirreza@gmail.com](mailto:naziriamirreza@gmail.com)

## SUMMARY

Machine Learning Engineer with 3+ years of experience specializing in AI-driven applications using LLMs, Agentic AI, generative models, and foundation models. Proficient in Python, PyTorch, and TensorFlow, with hands-on experience building and deploying novel models for AI-driven drug discovery.

## WORK EXPERIENCE

### Machine Learning Affiliate Researcher, Vector Institute

Jan 2024 - Dec 2025

- Developed and fine-tuned generative models (Diffusion, VAEs) and LLMs for predictive analysis in AI-driven drug discovery, working with molecular representations (SMILES) and genomic data (scRNA-seq).
- Applied Graph Neural Networks (GNNs) and foundation models (scFoundation, scGPT, Geneformer, and scBERT) to high-dimensional scRNA-seq data to improve the accuracy and speed of biological data interpretation.
- Implemented Bayesian optimization and uncertainty quantification techniques to improve model accuracy and interpretability for predictive modeling.
- Engineered and prototyped scalable ML reference applications, translating novel AI research into real-world solutions for commercialization pathways.

### Machine Learning Instructor and Mentor, AI4Good Lab (Mila)

May 2025 - June 2025

- Mentored a cohort of five students in a program empowering women and non-binary people in AI, guiding them to develop ML solutions for social-impact challenges.
- Designed and delivered weekly lectures and live-coding workshops covering data ethics, generative models, and production deployment.
- Guided the cohort through the full project lifecycle, from problem scoping and dataset curation to model validation and stakeholder presentations.

### Machine Learning Researcher, Amirkabir University of Technology

Sep 2022 - Sep 2023

- Fine-tuned BERT models for automated misspelling correction, achieving reduction in error rates and enhancing text accuracy.
- Developed and deployed a full-stack Django web application to showcase the model's AI capabilities, enhancing its accessibility and usability.
- Contributed to open-source NLP libraries, implementing new features that enhanced text-processing capabilities for underrepresented languages.

### Machine Learning Engineer, Sharif DeepMine

July 2021 - Sep 2021

- Developed and optimized text normalization pipelines using Python, NLTK, and SpaCy to address complex linguistic challenges in Persian text processing.
- Conducted a comparative analysis of open-source NLP tools to evaluate and recommend the optimal technology stack for the company's data pipeline.

## EDUCATION

2024 - 2026 M.Sc (Computer Science) at **York University**

(GPA: 4.0/4.0)

2018 - 2023 Bachelor's Degree at **Amirkabir University of Technology**

(GPA: 18.94/20.0)

## PROJECTS

---

### Diabetes Prediction using XGBoost

[Source Code on GitHub](#)

Developed an XGBoost model to predict diabetes from clinical data, achieving 75% prediction accuracy. Engineered the full data pipeline for preprocessing, feature analysis, and visualization using Pandas and Matplotlib.

### Real-Time Twitter Data Sentiment Analysis

[Source Code on GitHub](#)

Built a real-time sentiment analysis pipeline using Spark ML (Logistic Regression) to classify live-streaming tweets. Integrated Spark SQL Streaming for continuous data ingestion and processing for dynamic trend analysis.

### Additional Projects

[View More on GitHub](#)

Explore additional machine learning and generative AI projects, including LLM fine-tuning, diffusion modeling, and biomedical data analysis.

## PUBLICATIONS

---

Naziri, Amirreza and Hossein Zeinali (2023). “A Comprehensive Approach to Misspelling Correction with BERT and Levenshtein Distance”. URL: <https://arxiv.org/abs/2407.17383>.

Naziri, Amirreza, Arash Asgari, Aijun An, et al. (2025). “From Bias to Breakdown: Benchmarking Failure Mode Analysis of Single-cell RNA Sequencing Foundation Models in Acute Myeloid Leukemia”. In: *Proceedings of the AAAI Symposium Series*. Vol. 7. 1, pp. 553–557.

Naziri, Amirreza, Arash Asgari, Eleftherios Sachlos, et al. (2025). “Improving Classification of Cell Types in Acute Myeloid Leukemia with Self-guided Masking Technique”. In: *NeurIPS 2025 Workshop on AI Virtual Cells and Instruments: A New Era in Drug Discovery and Development*. NeurIPS, AI4D3.

## SKILLS

---

Languages	Python, R, JavaScript, SQL
ML & AI Frameworks	PyTorch, TensorFlow (Keras), Scikit-learn, HuggingFace, LLM fine-tuning
LLM Engineering	vLLM, HuggingFace, n8n, Ollama, Agentic AI, Retrieval-Augmented Generation (RAGs)
Data Platforms	Spark (PySpark), Hadoop (MapReduce), Hive, SLURM
Data Analysis & Vis	Pandas, NumPy, Matplotlib, Seaborn, PowerBI (DAX)
MLOps & Cloud	Docker, Kubernetes, AWS, Linux, Bash Scripting
Databases	MySQL, PostgreSQL, MongoDB
Web Frameworks	Django, Flask, FastAPI

## VOLUNTEERING

---

### Trainee Representative, Knowledge Mobilization, Connected Minds

Oct 2024 - Sep 2025

- Represented the trainee cohort on the Knowledge Mobilization steering committee, translating trainee needs into actionable program feedback.
- Helped the committee develop strategies to translate complex AI research into accessible insights for industry partners, policymakers, and the public.

## ACHIEVEMENTS

---

### Interdisciplinary AI Scholarship

2023-2025

- Awarded \$20,000 for groundbreaking contributions in AI research, particularly generative modeling and AI applications in healthcare.