

# Arian Naseh

Toronto, ON — +1-647-561-2559 — a.nasehzade@gmail.com  
Personal Website — LinkedIn — Medium Blog

## Summary

---

Applied Machine Learning Engineer and Researcher with 5 years of industry experience designing and deploying scalable models for retrieval, ranking, and predictive analytics. Combines an academic foundation in advanced ML algorithms with hands-on expertise in translating complex research into production-grade, high-throughput data pipelines. Proven track record of improving operational efficiency, driving automation, and utilizing large-scale machine learning to deliver scientifically robust, customer-centric solutions.

## Core Competencies

---

|                                |                                                                                                                       |
|--------------------------------|-----------------------------------------------------------------------------------------------------------------------|
| <b>Programming &amp; Data:</b> | Python (Primary), Java, SQL, Shell Scripting, Snowflake, BigQuery, Blob Storage                                       |
| <b>ML &amp; Modeling:</b>      | Classification, Regression, Ranking, XGBoost, CatBoost, LightGBM, PyTorch, TensorFlow, Time-Series Forecasting        |
| <b>Search &amp; Retrieval:</b> | Retrieval-Augmented Generation (RAG), Vector/Hybrid Search (Azure AI Search), Azure OpenAI Embeddings, ChromaDB, DSPy |
| <b>ML Infrastructure:</b>      | Online Experimentation, Azure ML, GCP, CI/CD, Drift Detection, Docker, FastAPI, Asynchronous Queuing                  |

## Professional Experience

---

|                                         |                    |
|-----------------------------------------|--------------------|
| <b>Northbridge Financial</b>            | Toronto, ON        |
| <i>Senior Machine Learning Engineer</i> | May 2022 – Present |

### Search, Retrieval & Intelligent Automation

- Engineered a RAG-based automated quote generation pipeline for heterogeneous broker submissions, projected to reduce time-to-quote from 1-2 days to under 5 minutes.
- Designed document-aware chunking algorithms mapping unstructured PDFs to Markdown hierarchies, ensuring semantic cohesion and high-fidelity retrieval.
- Integrated Azure AI Search and OpenAI embeddings with downstream rating engines, a workflow anticipated to increase quote-to-submission ratios by 3X.
- Architected structured LLM data extraction via Pydantic validation, establishing a scalable system expected to boost bind-to-quote ratios and drive top-line revenue growth.
- Automated >70% of First Notice of Loss (FNOL) intake via a LangChain and FastAPI microservice, significantly reducing manual processing bottlenecks.
- Engineered an asynchronous queuing mechanism to manage Azure OpenAI token limits, ensuring low-latency inference for high-concurrency request workloads.

### Ranking, Predictive Modeling & ML Systems

- Deployed XGBoost frequency and severity models for targeted risk interventions, yielding over \$1M in annual operational savings.
- Designed organization-wide AI monitoring dashboards to evaluate production automation rates, extraction accuracy, and model failure modes.
- Operationalized automated retraining and drift monitoring pipelines, proactively preventing silent production degradation and ensuring long-term model stability.
- Built MLForecast and Prophet time-series models with quarterly retraining, securing a \$500K resource expansion by directly informing executive workload strategy.

- Developed robust ETL pipelines utilizing Snowflake and AzureML to integrate and transform large-scale datasets for enterprise ML applications.

## Docma

Toronto, ON

*Machine Learning Consultant (Part-Time Contract)*

June 2023 – Jan 2024

- Optimized high-throughput API endpoints for advertising analytics, significantly reducing latency for real-time campaign dashboards.
- Refactored audience segmentation models, improving execution efficiency while preserving rigorous statistical fidelity and targeting accuracy.

## TicTie Labs

Toronto, ON

*Data Scientist*

June 2021 – May 2022

- Engineered supervised machine learning algorithms to predict yield, successfully optimizing environmental controls for indoor farming facilities.
- Developed real-time SQL and Python data pipelines to track operational metrics, driving data-informed agricultural decisions.

## York University

Toronto, ON

*Research Assistant*

Sept 2019 – May 2021

- Applied Deep Reinforcement Learning algorithms to optimize IoT network caching strategies, effectively reducing systemic energy consumption by 30%.

## Education

---

### York University

Toronto, ON

M.Sc. Computer Engineering (GPA: A)

Sept 2019 – May 2021

*Key Courses: Machine Learning Theory, Probabilistic Models and Machine Learning, Data Mining*

### Amirkabir University of Technology

Tehran, Iran

M.Sc. Digital Systems – Electronics (GPA: A)

Sept 2016 – May 2019

*Key Courses: Statistical Machine Learning, Artificial Intelligence, Computer Vision, Digital Signal Processing*

### University of Guilan

Rasht, Iran

B.Sc. Electrical Engineering – Electronics (GPA: B+)

Sept 2011 – Apr 2016

*Key Courses: Statistics, Algebra, Signal and System Analysis*

## Publications

---

- A. Nasehzadeh and P. Wang, “A Deep Reinforcement Learning-Based Caching Strategy for Internet of Things,” *IEEE/CIC ICC*, 2020.
- H. Wu, A. Nasehzadeh and P. Wang, “A Deep Reinforcement Learning-Based Caching Strategy for IoT Networks With Transient Data,” *IEEE Transactions on Vehicular Technology*, 2022.