Sourena Khanzadeh

AI Research Engineer

LLM Agents | Evaluation | Tool Use | Deep Learning | ML Systems

Toronto, Ontario, Canada | 416-456-0091| sourena.khanzadeh@gmail.com

sourena-khanzadeh.com | github.com/skhanzad | linkedin.com/in/sourenaK | Google Scholar

Professional Summary

AI Research Engineer with a Ph.D. in Computer Science and 3+ years of experience building and evaluating LLM agents, deep-learning pipelines, and search-based AI systems across industry, government, and academia. Develops agent harnesses, retrieval and tool-use workflows, evaluation infrastructure, and reproducible ML prototypes; published at AAAI, Canadian AI, and in peer-reviewed journals.

Technical Skills

Languages Python, C/C++, TypeScript, Java
Machine Learning PyTorch, Hugging Face, scikit-learn, XGBoost, deep learning, reinforcement learning, fine-tuning, GANs, synthetic data
LLM & Agent Systems LLM agents, tool and function calling, multi-agent orchestration, retrieval-augmented generation (RAG), context and memory management
Evaluation Agent harnesses, trajectory logging, counterfactual evaluation, automated scoring, experiment design, observability, guardrails, failure analysis
Infrastructure & Software Docker, Kubernetes, AWS, PostgreSQL, Redis, Git, GitHub Actions, APIs, testing, reproducible pipelines

Professional Experience

AI Engineer, Agentic Systems **May 2026 – Present**
Flybits | Full-time | Supervisor: Dr. Hossein Rahnama *Toronto*

- Build agentic AI systems that plan, call tools across internal and partner APIs, and complete multi-step tasks, with every decision logged for end-to-end auditability.
- Implement retrieval over structured user-context graphs and pair learned models with symbolic constraints, so generated responses stay grounded in real customer state instead of model priors.
- Stand up the evaluation loop — task completion, groundedness, faithfulness, recovery from tool failure — and use it to tune model routing across capability, latency, and cost.

AI Research Engineer, sAIpien **May 2026 – Present**
MIT Media Lab | Part-time research affiliation | Supervisor: Dr. Hossein Rahnama

- Prototype Perspective-Aware agents, privacy-preserving long-term memory through adaptive knowledge graphs (*Chronicles*), and HCI^2 benchmarks for transparency and human oversight.

Postdoctoral Fellow / AI Research Engineer **May 2026 – Present**
Toronto Metropolitan University | Part-time academic appointment *Toronto*

- Build evaluation harnesses in Python and PyTorch — trajectory logging, counterfactual replay, automated scoring — that measure LLM-agent reasoning, orchestration, context management, and reliability.
- Take research questions to working prototypes with experimental protocols, metrics, and publications, shipping code that other researchers can run and reproduce.

Machine Learning Engineer (Graduate Research) **Jan 2023 – Apr 2026**
Toronto Metropolitan University *Toronto*

- Built AI prototypes spanning multi-agent systems, heuristic search and planning, deep learning, reinforcement learning, and GAN-based synthetic-data generation.
- Implemented training and inference pipelines with PyTorch, Hugging Face, scikit-learn, and XGBoost; containerized experiments with Docker and automated checks through GitHub Actions CI.
- Ran controlled experiments with ablations, profiled failure modes, and published the results at AAAI, Canadian AI, and in peer-reviewed journals.

ML Engineer / Research Scientist Intern **Jan 2024 – Jan 2025**
National Research Council Canada *Toronto*

- Prototyped and evaluated knowledge-informed machine-learning models for anomaly detection over large-scale, severely imbalanced telemetry data.
- Engineered domain-derived features that reduced dependence on large labelled datasets and supported reproducible, peer-reviewed experiments across an interdisciplinary team.

Selected Projects

Project Ariadne — Faithfulness Auditing for LLM Agents **2026**

- Built an evaluation harness that captures agent reasoning traces, applies hard counterfactual interventions, replays the workflow from the modified state, and scores whether the stated reasoning caused the output.
- Audited 30 agent trajectories and surfaced a 76.7% faithfulness-violation rate, rising to 96% on scientific reasoning — a reusable protocol for chain-of-thought monitoring and agent regression testing.

AgentMesh — Multi-Agent Coding System **2025**

- Engineered a Python framework orchestrating specialized Planner, Coder, Debugger, and Reviewer agents through an iterative loop with execution feedback and automatic self-repair.
- Converted high-level requirements end to end into non-trivial, tested software with minimal human intervention.

GANsemble — Synthetic Data Generation Pipeline **Canadian AI 2024**

- Combined factorial augmentation search, transfer learning, and a class-conditional GAN to automate synthetic-data generation for a 210-image imbalanced dataset.
- Lifted downstream classification accuracy to 91.5% — 4 points over duplication-based oversampling and 5.5 points over no oversampling.

Search-Based Planning and Automated Code Improvement **2023–2026**

- Co-developed a restarting-random-walk planner, published at AAAI 2026, that was faster or uniquely successful on 70.3% of unbounded-region tasks while using linear rather than exponential worst-case memory — restart policies that carry over to test-time search.
- Built admissible search heuristics for automated refactoring, cutting node expansions by 82% and runtime by 74% against uninformed search.

Education

Ph.D. in Computer Science **Sep 2023 – May 2026**
Toronto Metropolitan University

Research focus: Agentic AI workflows for reliable automated reasoning — LLM-agent orchestration, reasoning loops, tool use, and evaluation of agent faithfulness. **GPA:** A+. Coursework: Heuristic Search, Deep Learning, Directed Intelligent Robotic Systems.

Bachelor of Computer Science **Sep 2018 – Dec 2022**
Toronto Metropolitan University

A+ in every AI course: Machine Learning, Artificial Intelligence, Reinforcement Learning, and Computer Vision.