

Khashayar Bayati

AI Engineer | Agentic Systems, Forecasting, Evaluation
Toronto, ON | khashayarbayati1@gmail.com | khashayarbayati.com
[linkedin.com/in/khashayarbayati](https://www.linkedin.com/in/khashayarbayati) | github.com/Khashayarbayati1

SUMMARY

AI engineer with a PhD in Electrical and Computer Engineering, building production AI systems end to end at Unilever's AI lab in Toronto: an agentic optimization platform used by procurement buyers, ownership of three production optimizers, time-series foundation model benchmarking, and the evaluation that decides whether an LLM agent is ready for daily use. Sole builder of Kaivest, a live personal finance platform with an AI assistant over real bank data. Peer-reviewed publications in machine learning.

SKILLS

GenAI and Agentic Systems: Conversational AI, agent orchestration (LangGraph, LangChain, OpenAI Agents SDK, Gemini function calling), tool calling and structured outputs, multi-agent workflows, human-in-the-loop review and approval patterns, prompt engineering and context strategies, LoRA fine-tuning, Hugging Face Transformers

RAG and Retrieval: Embeddings and chunking, vector retrieval (Chroma), relevance ranking and LLM relevancy scoring, response grounding, end-to-end validation of retrieved methods on standardized benchmarks, benchmark dataset design

Evaluation and Observability: LLM-as-judge scoring (hallucination/faithfulness, completeness), regression suites with pass@k reliability scoring, production prompt replay, per-request trace, token, and cost tracing on an internal platform comparable to LangSmith/Langfuse, leakage-aware benchmarking

Machine Learning: PyTorch, TensorFlow, time-series forecasting (foundation models to classical), Bayesian and probabilistic modeling, causal inference, experimental design

Engineering: Python, TypeScript, SQL, Git, FastAPI, Streamlit, Docker, GCP, Azure DevOps, GPU computing with SLURM, CI-gated testing (Vitest, Playwright)

PROFESSIONAL EXPERIENCE

Data Scientist | Unilever, Horizon3 AI Labs, Toronto Jan 2026 - Present

- Lead development of an agentic platform (Python, LangGraph) for multi-hop distribution optimization, used by the procurement team behind a multi-billion-dollar supply chain; buyers run optimizations, interpret results, and diagnose infeasibility conversationally, with a human-in-the-loop step for adjusting constraints before solver runs, replacing repeated spreadsheet exchanges across teams. Runs on the lab's on-prem GPU server.
- Own evaluation and observability for a retail-facing conversational product agent (Python, FastAPI, Gemini function calling) built to connect with retailer shopping agents.
- Took over three production optimizers from an outgoing vendor, including the downstream multi-hop distribution optimizer: ran each tool locally, documented the codebases, and remain the point of contact when the business hits an issue.
- Instrumented the agent on an internal LLM observability platform, comparable to LangSmith/Langfuse, with per-request trace, token, and cost visibility plus LLM-as-judge scoring for hallucination (faithfulness) and completeness; built a 41-case regression suite with pass@k reliability scoring and used it to benchmark a major architecture change against its quality-versus-latency trade-off.
- Replayed production prompt data through the traced agent and surfaced a 25% failure rate on real customer questions that the existing test suite had missed; diagnosed and fixed a model-migration break in the tool-calling layer.
- Designed the retrieval, ranking, and evaluation architecture for a research agent over arXiv, Semantic Scholar, and GitHub: Chroma vector store, vector search and relevance ranking, and a human-in-the-loop step where users select sources from LLM-generated relevance reports; closed the loop by implementing retrieved methods and benchmarking them on standardized datasets (precision, recall, F1), verifying the agent's selections fit the task.
- Built a mathematical infeasibility-analysis capability for a constrained supply-chain optimizer, built to support Q3 2026 planning; cuts root-cause diagnosis from days or weeks of analyst work to minutes and points buyers to the exact inputs to change.
- Built six-month commodity price forecasting models with exogenous market drivers, reaching roughly 9% test MAPE across six product grades; gathered requirements from and presented results to buyers, product owners, and technical stakeholders.

- Benchmark eight time-series foundation model families across three commodities with a leakage-aware evaluation protocol; implemented LoRA fine-tuning of TimesFM and Chronos, presented the methodology at a Vector Institute Learn Day, and mentor two University of Toronto MScAC interns.
- Built, during the Vector Institute Agentic Forecasting Bootcamp (cohort included teams from TD, BMO, and Telus), a news-driven multi-agent LLM system for probabilistic price forecasting that ranked first of 14 methods on mean CRPS across 1 to 13 week horizons, with a skill score of +0.27 over the naive baseline.

PhD Researcher, Machine Learning | Toronto Metropolitan University *Sep 2020 - May 2025*

- Designed a variational autoencoder and Transformer framework for multivariate time-series modeling and causal discovery, improving F1 score by 18% and reducing false positives by 12% over baselines; published in Knowledge-Based Systems.
- Built reusable Bayesian inference and experimentation pipelines in Python and PyTorch, training and evaluating models on Digital Research Alliance of Canada GPU clusters through SLURM.
- Course instructor for Signals and Systems and Random Processes; received a Dean's Teaching Award and mentored junior graduate students on ML implementation and experiment design.

AI Research Intern | Vector Institute, Toronto *Sep 2023 - Dec 2023*

- Implemented and benchmarked HMC, SVGD, MALA, and ULA sampling algorithms for deep generative models on Vector's Mars and Vaughan GPU clusters, improving convergence stability by 25% and uncertainty calibration by 18%.

Signal Processing Engineer | Oxymap *Mar 2019 - Nov 2020*

- Designed signal processing algorithms combining stochastic noise modeling and statistical filtering, improving accuracy by about 20% in physiological signal estimation; worked with cross-functional teams to move research-grade models into production.

Research Assistant | University of Tehran *Sep 2017 - Aug 2020*

- Trained supervised models for time-series classification, reaching 94% accuracy across 12 subjects, benchmarked with cross-validation and permutation testing.

SELECTED PROJECT

Kaivest, AI-Powered Personal Finance Platform | kaivest.ai, sole builder *May 2026 - Present*

- Built and shipped a live personal finance platform managed through conversation, with web (Next.js, TypeScript) and mobile (React Native, Expo) clients, Plaid bank linking, Stripe billing, and a Supabase Postgres backend with row-level security; invite-only beta with 15 users.
- Developed the agentic assistant on Anthropic Claude via the Vercel AI SDK, exposing about 60 tools across 14 modules with automatic routing between models based on query complexity.
- Built an evaluation harness that benchmarks candidate models across the assistant's full tool surface and used it to set production model routing; the roughly 57,000-line TypeScript codebase carries about 1,200 unit tests, 49 end-to-end tests, CI checks gating every merge, and Sentry error tracking.

EDUCATION

- PhD, Electrical and Computer Engineering, Toronto Metropolitan University, 2025. Machine learning, causal inference, and time-series modeling.
- MSc, Electrical Engineering, University of Tehran, 2020.
- BSc, Electrical Engineering, University of Tehran, 2016.

PUBLICATIONS

- DVAE: A Dynamic Variational Autoencoder for Structured Causal Discovery with Application in Biomedical Time Series. Knowledge-Based Systems, vol. 335, 2026.
- Reliable Truncation Parameter Selection and Model Order Estimation for Stochastic Subspace Identification. Journal of the Franklin Institute, vol. 362, no. 11, 2025.
- Granger Causality Pattern Learning Equipped with Noise Invalidation Soft Thresholding. EUSIPCO 2023, IEEE.

CERTIFICATIONS

- SQL, HackerRank, 2025 | Machine Learning in Production (MLOps), DeepLearning.AI, 2025