Adam Kuca

Technical founder and AI research engineer building agent systems and researching self-improvement.

Ann Arbor, MI — (734) 773-7019 — adamkuca@umich.edu — adamkuca.vercel.app
github.com/GSXRcoder159 — linkedin.com/in/adam-kuca-505409286

Education

University of Michigan, Ann Arbor, MI *Expected Dec 2026*
M.S. in Computer Science & Engineering

- GPA 4.0. Focus: AI agents, reinforcement learning, large language models, and natural language processing.

University of St Andrews, Scotland *Graduated 2025*
B.Sc. (Hons) Computer Science — First Class Honours (4.0 GPA)

- Dean's List every year (2022–2025). Completed degree a year early.

Industry & Startup Experience

Transmute — AI Capability Compiler *Aug 2026–Present*

- Exploring how agent traces and outcomes can become specialized software: deterministic tools for repeated work, small models for judgment, and frontier-model fallback for unfamiliar cases.
- Built a hand-specialized retail-agent prototype using a small model and deterministic code. On 180 generated held-out τ^2-bench tasks, single-attempt success was 93.9% versus GPT-5.2's 94.3%, at approximately 18× lower inference cost. Two of three runs failed stricter acceptance gates.

MAMA AI *May–Jul 2026*
AI Research Engineer Intern Prague, Czechia

- Built an evaluation and optimization framework for production LLM pipelines from scratch, with labeled datasets, class-weighted metrics, checkpointing, and parallel execution on Azure and local models.
- Compared MIPROv2, GEPA, TextGrad, ensembles, and prompt merging on two production pipelines. Increased extraction accuracy on multimodal documents (text, PDF, scanned images) from 77% to 87%.

Insora — Secure AI Workspace for SMEs *Jul 2025–Jul 2026*
Co-Founder & CTO insora.dev

- Designed and built a production AI workspace that breaks down important decisions into testable hypotheses and grounded evidence.
- Built hybrid graph+vector retrieval (Neo4j, Qdrant), retrieval and verification agents, and a full-stack workspace (Vite, FastAPI, PostgreSQL, Redis, Celery), deployed on Hetzner and Azure.
- Led 50+ discovery conversations across about 25 EU SMEs and delivered a client pilot in six days. The pilot answered 25 expert questions across 150+ scientific papers with grounded citations and no hallucinations or contradictions in the pilot evaluation.

Blindspot AI *Jul–Aug 2023*
Machine Learning Engineer Intern — AskYourData Prague, Czechia

- Improved enterprise RAG retrieval accuracy by 70%+ through prompt, embedding, chunking, and index tuning; built real-time web-search and ingestion pipeline in Scala and Azure Functions.

Research

Agent Harnesses for Cloud Infrastructure *Sep 2026–Present*
University of Michigan research; two-person project with Kai Kin

- Studied xAI's Grok Build and built independent evaluation using Terraform validation, offline plans with mocked AWS, and OPA policies. Built a 180-line Python ReAct harness that raised Qwen 27B validation from 26% to 98% on 50 tasks (one run per task); Qwen 9B reached 90% versus Grok Build's 78%, at about a quarter to a third of the tokens.
- Added searchable Terraform documentation, plan summaries, and requirement checks. On 50 held-out tasks, Qwen 27B policy pass rate rose from 18% to 31% and Claude-judged score from 0.45 to 0.58, matching Qwen Max on both. Now testing multi-file projects.

Verified Self-Play for Compiler Optimization *Sep 2026–Present*
Research with Gabriel Poesia, University of Michigan Token Games

- Built a Python duel harness and C++ LLVM slicer for 69 MB of IR from 10 codebases. Combined alive2 formal verification, differential tests, and paired timing, reducing timing noise from 28% to less than 0.7%.
- Found verified optimizations to two fragments from Google's re2 that ran 14–16% faster on real hardware than LLVM's -O3. Caught invalid compiler flags that passed native tests and instruction-count improvements that slowed real code. Self-play did not beat the single-model baseline on 50 fresh fragments; now exploring PyTorch optimization.

Project Vera — World Models for Cloud Infrastructure *2026–Present*
Research with Prof. Ang Chen, University of Michigan Website

- Working on world-model emulators for cloud-infrastructure agents and execution environments for testing and training them.

Automated Constraint Model Modification with LLMs *Jan–Aug 2025*
Research Intern — University of St Andrews GitHub

- Defined task and taxonomy for editing formal constraint-programming models from natural-language requests; created benchmark and implemented pipeline achieving 95.8% compilation accuracy with GPT-4o. Presented as first-author paper at LLMSolve Workshop, CP 2025 (Glasgow).

Selected Projects

Parity — Stripe Billing Simulator for AI Agents *Sep 2026–Present*
Independent project; built with AI coding agents

- Built a stateful TypeScript simulator in one week: subscriptions, proration, invoices, signed webhooks, and test clocks. It works as both an MCP server and an RL environment, so agents can be tested without using the real API.
- Built a nightly benchmark against real Stripe with status, shape, value, and sequence-level checks. Held-out value fidelity rose from 8.8% to 39.1% through automatic and LLM-assisted patches; an agent succeeded or failed the same way on Parity and real Stripe in 79/80 billing tasks.

Personal AI Assistant — Self-Improving Agent System *Sep 2026–Present*
Independent project; running daily for personal use

- Designed and directed the build of a Rust/TypeScript assistant using AI coding agents. The assistant can change its own tools and code. Verified live updates with conversation continuity and automatic rollback; a compiled kernel enforces approvals, usage limits, and protected paths. The assistant added its own diagnostic tool and read-only Gmail and Calendar integrations.

Slim — Delete Your Dependencies *Aug–Sep 2026*
Open-source TypeScript CLI and GitHub Actions GitHub

- Built a TypeScript CLI that finds how a repo uses an npm dependency, generates a smaller replacement, and differentially fuzzes it against the original. Opens a PR with evidence and tests; includes an LLM generator, node:test/Vitest trace hooks, and GitHub Actions. CI runs on Ubuntu, macOS, and Windows.

Entrepreneurship & Recognition

Selected for YC Startup School (2026) and Radical Ventures AI Founders Masterclass (2025); EF London finalist (2025). Michigan Entrepreneurial Leadership Program (2026). Merged CAMEL-AI contribution.

Selected Technical Skills

Languages & Systems: Python, TypeScript, Rust, C/C++, Scala; FastAPI, React, Next.js; PostgreSQL, Redis, Neo4j, Qdrant; Docker, Terraform, AWS, Azure, GitHub Actions.

AI & Research: PyTorch, Hugging Face, vLLM, TorchRL; agent harnesses, retrieval, self-play, DPO/KTO, PPO/MAPPO; DSPy, GEPA, TextGrad; LLVM, alive2, OPA, differential testing.