Kevin Saylor Ruschman

AI Systems Architect · Agent Infrastructure, Multi-GPU Clusters, Post-Training & Full-Stack Systems

📍 Los Angeles, CA • 📞 (562) 713-2084 • ✉️ ksruschman@gmail.com • 🌐 krusch.dev • 💻 github.com/kruschdev • 🤗 hf.co/krush420 • 👔 LinkedIn

EXECUTIVE SUMMARY

AI Infrastructure Engineer, Backend/Full-Stack Engineer, and Open-Source Maintainer building the reliability infrastructure that makes autonomous AI systems safe, persistent, and useful—evaluated across 26 whitepapers and 335,000+ benchmark traces (95/95 SHA-256 gates). Creator of **krusch-stack v1.1** (39-tool MCP + Grounded [90.0% DABstep @ $0.81, 2.54x faster] & 2PC harnesses), **@krusch/toolkit v1.8.1** (39 Lego bricks, 16 static/AST gates, 104/104 compliance), **krusch-context-mcp v2.0** (82★ — **90.26% Snap LoCoMo** [95.77% recall @ $0.00 ingest], **91.17%** [547/600 @ $24.54, uploaded `dolphinbench.ai` submission queued awaiting acceptance; > 70.67% published entry], 90.67% Grounded v0.3 [90.50% Riley], 81.50% Luna @ $1.35 [71x < Mem0], **57.33%** top open-weights self-run result @ $0.00 across 28 runs [98.33% Oracle@28], 2K/2K BEAM), **krusch-cascade-router** (<15µs Stage-0 CPU gate — 0.7263 CI-verified RouterArena PR #169), and **krusch-coding-harness** (0/1,046 AgentDojo ASR; 51.56% Polyglot). Eagle Scout.

CORE TECHNICAL CAPABILITIES

Agent Infrastructure & Harnesses: Node.js 22 ESM, Python 3.12, `@krusch/toolkit v1.8.1` (39 bricks, 16 Slop-Breaker rules), MCP (`krusch-stack v1.1` 39 tools, Grounded v0.3 & 2PC harnesses), `SovereignMemoryStore` (`0o444` SQLite), 0.00% ASR Policy Floors (`bwrap`), <15µs Stage-0 CPU Cascade Routing.

Relational State & Full-Stack Web: PostgreSQL 16 State Planes, `pgvector` (HNSW), Hybrid FTS5/BM25 + Dense RRF, SQLite WAL, DBOS Durable Execution, FastAPI, SQLAlchemy 2.0, Express, React, Vite, Tailwind CSS, Cloudflare Pages & Workers.

Post-Training, Evals & GPU Inference: Wave-4–6 QLoRA+DPO (31B 57.33% open-weights, 120B MXFP4, 30B MoE 43.00% & 8B 15.7x lift on 96GB G4; 18.1K SFT+DPO rows), Hermetic Evals (90.26% Snap LoCoMo @ $0.00 ingest; 28-run DolphinBench 91.17% queued / 90.67% v0.3 / 98.33% Oracle@28; 90.0% DABstep; 2K/2K BEAM), `snowflake-arctic-embed2` .

Bare-Metal Clusters & Document Engines: 8-Node Bare-Metal Linux Orchestration (Dual Xeon 40T, 57GB+ Local VRAM + 96GB G4 Burst, 10GbE Fabric, Tailscale), Layout-True PDF Spines (Poppler TSV, 7 Parser Families), Statutory/Contract Precedence DAGs, Cryptographic Purge.

ENGINEERING EXPERIENCE & OPEN-SOURCE SYSTEMS

AI Infrastructure Engineer & Open-Source Maintainer · Independent R&D Lab (@kruschdev)

2023 – Present · Los Angeles, CA

- **Unified 7-Pillar MCP Stack & Zero-LLM Persistent Memory (krusch-stack v1.1 & krusch-context-mcp v2.0):** Authored open-source `npx krusch-stack init` (39 MCP tools, `-68.5%` BM25 schema bytes, `krusch-grounded-harness` hitting **90.0%** [9/10] on Adyen `DABstep` @ $0.81 [`-74.5%` cost, `2.54x` faster], plus `INV-STACK-10` 2PC cutting dup calls `-95.2%` [`42→2`]), `krusch-git` , `krusch-fs-mcp` , and `krusch-context-mcp` (82★). Evaluated unmodified `$0.00` zero-LLM `0o444` `SovereignMemoryStore` across **Snap LoCoMo** (ACL 2024, 1,540 probes: **90.26%** acc / **95.77%** `dia_id` recall, 89.68% tight @ 4.5K tok, 92.66% Oracle@2), 28 `DolphinBench` 600-task runs (16,800 traces: **91.17%** [547/600 @ $24.54, uploaded submission `UUID c047b063` queued on `dolphinbench.ai` , > 70.67% published entry], **90.67%** [90.50% Riley] `Argon Low + Grounded v0.3` , 98.33% Oracle@28, **81.50%** `Luna + Grounded` @ $1.35 [71x < Mem0], and **57.33%** top open-weights Wave-6 `Gemma 4 31B` @ $0.00 [54.17% queued stock]), and 2,000/2,000 BEAM (93.00% 10M IE).
- **Transactional 2PC Coding Harness & Policy Floor (krusch-coding-harness):** Engineered an open-source coding harness enforcing import-frozen policy floors, shell-deobfuscation checks, truncation-poisoning breakers, verify-on-stop outer loops, and Bubblewrap (`bwrap`) sandboxes committing verified diffs via a PostgreSQL write-ahead apply journal paired with POSIX atomic `renameSync` / `fsync` swaps and pre-commit SHA-256 drift guards (<32ms local journal+swap overhead)—achieving **0.00% ASR** across 1,046 ETH SpyLab AgentDojo security episodes and **51.56%** on Aider Polyglot (2.64x lift).
- **Autonomous SRE, Wave-4–6 QLoRA+DPO & Sub-15µs Cascade Routing (jean-sre, @krusch/toolkit v1.8.1, krusch-cascade-router):** Authored 39-brick `@krusch/toolkit v1.8.1` & `krusch_toolkit v1.6` (7 blueprints, 16 <1ms Slop-Breaker detectors, 104/104 compliance) and a 3-Family Referee pipeline publishing 18,131 zero-leak SFT+DPO rows (`krush420/sovereign-code-audit-sft`) to train 4 open-weights adapters—31B Wave-6 (**57.33%** top open-weights DolphinBench self-run result), 120B MXFP4 (39.50%), 30B MoE (43.00%), and 8B Edge (18.33%, **15.7x** lift)—paired with <15µs pure-CPU Stage-0 cascade routing prior to model inference (0.7263 CI-verified RouterArena PR #169 @ $0.38/1K, 26.4x < GPT-5; top Pareto on LLMRouterBench & RouterBench self-run evaluations) and 26 technical whitepapers.
- **Multi-Family 5-GPU Cluster & 96GB Blackwell Burst Tier:** Designed and operate an 8-node bare-metal Linux lab (Dual Xeon 40T, 144GB+ ECC/DDR4 RAM, 57GB+ local VRAM across Dual RTX 3060 24GB & Dual RTX 2080 Ti 22GB NVLink over 10GbE + on-demand 96GB Blackwell G4 burst) pinning orthogonal model families (`Qwen3-30B-MoE` at 124.2 tok/s, `Gemma 4` , `Hermes 3` , `snowflake-arctic-embed2`) 100% in VRAM alongside PostgreSQL 16 + `pgvector` HNSW and DBOS queues at 99.9% uptime.
- **Offline-Capable Domain Reference Architectures (KruschNexus, KruschLaw & KruschBiz):** Built end-to-end reference engines demonstrating zero-egress PDF layout parsing (`KruschNexus` 7-parser Poppler TSV spine: 26.74% Macro F1 & 33.95% MAUD Recall@2 across 6,889 `LegalBench-RAG` queries, surpassing OpenAI+Cohere Recall@64), two-pass statutory proposition grounding & 16 temporal `as_of_date` Authority Packs (`KruschLaw` : 0.00% False Support Rate on labeled statutory golden sets, verifiable SHA-256 purge), and confirmed-edge contract precedence DAGs with draft isolation & statutory compliance joins (`KruschBiz`).

Technical Operations, Management & Applied Systems · Prior Experience

2014 – 2023

- **Applied Electrical & Mechanical Construction (Southern Vermont):** Executed electrical panel modifications, conduit installation, and structural carpentry, building hands-on mastery of power budgeting, grounding, and physical systems.
- **Operations Supervisor (Wisconsin Bicycle Federation) & Cafe Operations Partner (Aspen, CO):** Directed multi-person mobile diagnostic crews and field logistics; launched guest operations, POS systems, supply chains, and staff training to first-year profitability.

EDUCATION & DISTINCTIONS

Fullerton College — Associate of Arts (AA) — Graduated 2012 · Fullerton, CA

Eagle Scout & Senior Patrol Leader — Boy Scouts of America (Troop 883) — Earned 2010 · La Habra, CA

*Earned highest rank in scouting after 11 years of scouting leadership. Planned and executed a 100+ hour community service initiative directing a 9-person crew.