KAI STEPHENS

OPEN-SOURCE AI & SOFTWARE ENGINEER

Newmarket, Ontario | kai@kaios.ca | kaios.ca | GitHub | Hugging Face | LinkedIn

PROFILE

Open-source AI developer building language models, evaluation benchmarks, and agent infrastructure. Published 14 models and two datasets on Hugging Face; Carnice V2 crossed one million downloads.

EXPERIENCE

The Neutrality Project | *Co-Founder, AI Research & Engineering* **Jul 2026 - Present**

Remote

- Co-built an open behavioral benchmark spanning 3,987 public-opinion questions, 14,395 answer choices, and six anchored political dimensions; published profiles for 24 models.
- Built reproducible collection, scoring, refusal-analysis, and review workflows; helped launch the project to 5M+ post views and 2,200 followers in its first 24 hours.

Nipux Inc. | *Founder, AI Systems Engineer* **Jun 2025 - Present**

Newmarket, Ontario

- Built a local-first control plane for long-running AI agents, combining hardware detection, runtime and model selection, checkpointed runs, browser sessions, durable logs, and a local web interface.
- Designed a web-to-daemon architecture in which the local service owns runtime management, state, browser control, and orchestration while users review machine and model plans before installation.

Treefrog Inc. | *Software Developer* **Mar 2024 - Oct 2025**

Newmarket, Ontario

- Contributed to web software and custom digital solutions in a full-time developer role at a digital product and marketing agency.

SELECTED TECHNICAL WORK

Carnice & Grug Open Models | *Python, PyTorch, Transformers, PEFT/QLoRA, vLLM, GGUF*

- Published 14 Hugging Face models and two datasets, including Carnice agent models for terminal, file, browser, repository, debugging, and multi-step workflows; Carnice V2 passed 1M downloads.
- Fine-tuned Grug 12B on 5,740 license-verified compact-reasoning examples; on a documented 36-row local proxy, reduced generated tokens 69.8% while accuracy moved from 91.7% to 100%.

UWaterloo Campus Map | *Next.js, TypeScript, MapLibre, OpenStreetMap, Cloudflare Workers*

- Built a keyless, client-side campus navigation app with live UW ArcGIS data and optimal A* routing across a 38.6K-node, 87.2K-arc walking graph, including turn-by-turn and tunnel directions.
- Packed the connected routing graph to approximately 786 KB gzipped and verified A* results against Dijkstra, symmetry, and edge cases before static deployment.

EDUCATION

University of Waterloo **Sep 2026 - Jun 2031**

Bachelor of Science, Geospatial Data Science

SKILLS

Languages: Python, TypeScript/JavaScript, Bash | **AI:** PyTorch, Transformers, PEFT/LoRA, vLLM, Hugging Face, llama.cpp/GGUF |
Software: Next.js, FastAPI, Flask, MapLibre, OpenCV, Docker, Linux, Git, Cloudflare Workers

Additional: Summited Mount Kilimanjaro at 15; shares open AI work with an audience of 25K+ on X.