

BENJAMIN WARNER

Whitby, Canada

benwarner1@gmail.com | 289-404-1840

linkedin.com/in/ben-warner-565597257

SUMMARY

Builder and researcher focused on practical AI systems. Founded and shipped **Axis MCP**, a commercial Model Context Protocol platform that handles authentication, metered billing, and dynamic routing across multiple AI tool servers at production scale. Conducting original research through **Orbit**, an open-source project investigating how scaffolding techniques — structured decoding, retrieval-augmented generation, and rigorous delta-scoring evaluation — can push 1–8B-parameter local LLMs toward competitive coding agent performance. Experienced across the full stack: Cloudflare Workers, Next.js, FastAPI, Python, and LLM tooling.

TECHNICAL SKILLS

- **AI & LLM:** Model Context Protocol (MCP), FastMCP, Ollama, Claude API, OpenAI API, LLM tool use, RAG, structured/grammar-constrained decoding, prompt engineering, agent loop design
- **Languages:** TypeScript, Python, JavaScript, SQL, Java, HTML, CSS
- **Frameworks & Runtimes:** Next.js 14, FastAPI, Cloudflare Workers (Wrangler), Node.js, Angular, React
- **Infrastructure & Data:** Supabase (Postgres, PostgREST, RPC), Docker, Coolify, Railway, Cloudflare Tunnels, Stripe (metered billing), Git, Linux, Azure, Google Cloud Platform
- **Tooling:** Vitest, Microsoft Power Platform (Power BI, Power Automate), Excel (VBA)

WORK HISTORY

Computer Science Co-Op Student May 2026 – Present
Ontario Power Generation Darlington, ON

- Returned for a second co-op rotation, continuing contributions to the Nuclear modernization program.
- Applying experience with TypeScript, web tooling, and automation to internal engineering initiatives across the Darlington site.

Computer Science Co-Op Student Jan 2025 – Dec 2025
Ontario Power Generation Darlington, ON

- Contributed to the Nuclear modernization project, delivering internal tooling and dashboards used by engineering teams across the Darlington site.
- Built production web applications in TypeScript, Angular, HTML, and CSS; shipped features on live internal systems serving hundreds of users.
- Developed Power Platform (Power BI, Power Automate) automations that reduced manual reporting overhead for the team.
- Mastered multiple new codebases and toolchains across concurrent assignments in a regulated, high-accountability environment.

Server/Bartender May 2023 – Aug 2024
The Wismer House Port Elgin, ON

- Delivered high-volume table service in a fast-paced environment; developed strong communication, teamwork, and situational adaptability.

Floor Worker / Shift Supervisor 2019 – 2024
The Beer Store & Tim Hortons Brooklin, ON

- Promoted to shift supervisor at Tim Hortons; managed cash handling, team coordination, and customer escalations across both roles.

PROJECTS & RESEARCH

Axis MCP — *Founder & Lead Engineer* 2025 – Present
axis-mcp.live

- Founded and shipped a commercial MCP platform that provides hosted AI tool servers to LLM-powered applications, launching June 2026.
- Engineered a Cloudflare Worker gateway that validates Bearer tokens against Supabase, enforces free-tier rate limits (500 req/month), and dynamically routes requests to multiple Python MCP spokes — globally distributed, zero cold-start.

- Designed full metered billing infrastructure: Stripe meter events, Supabase usage logging, a Cloudflare Cron Worker for billing reconciliation, and a race-condition-free SQL RPC (`FOR UPDATE SKIP LOCKED`) to eliminate double-billing.
- Shipped a Next.js 14 storefront and user dashboard with Stripe Checkout, API key management (Free: 1 key; Pro: unlimited), and live usage metrics.
- Deployed and operate multiple FastAPI + FastMCP spoke servers across Coolify (Raspberry Pi) and Railway; services include stock news, GO Transit data, financial data, and web scraping.
- Built **AxisCLI** (`axis-mcp-cli`): an open-source TypeScript CLI for managing local MCP server registries — 12 commands, 39 passing Vitest tests, ready for npm publication.

Orbit — *Researcher & Author*

2025 – Present

- Designing and building an open-source local coding agent to rigorously study how far scaffolding alone can push small LLMs (1–8B parameters, Ollama backend) as practical coding assistants.
- Implemented a two-phase agent loop separating JSON tool-call decisions (Phase 1) from natural-language responses (Phase 2) to mitigate the simultaneous reasoning and formatting failures endemic to sub-8B models.
- Built a 4-strategy malformed JSON parser and a BM25 + subword tokenization codebase search index supporting CamelCase and snake_case decomposition for accurate retrieval in real codebases.
- Surveyed and synthesized ~30 arxiv papers to ground design decisions in evidence: RepoCoder iterative RAG achieves +12–13 EM on 350M–6B models (arXiv:2303.12570); grammar-constrained decoding reaches near-100% schema compliance at 1–8B (arXiv:2501.10868); plain zero-shot CoT is *flat or negative* for small coding models (arXiv:2603.03406).
- Designing a 100-task delta-scoring evaluation harness (scaffolded – naked pass@1) with McNemar significance testing and explicit naked baseline protocol to measure scaffolding gains honestly and produce publishable results.

Additional Projects

- **Personal AI Assistant:** Voice-controlled assistant built in Python, extended with OpenAI API integration for natural language understanding.
- **Arduino & Robotics:** Built sensor-oriented robots and hardware–software systems using Arduino and open-source components.
- **Canadian Computing Competition (CCC):** Competed twice in high school, solving algorithmic problems under strict time constraints.

EDUCATION

Bachelor of Science: Computer Science

Ontario Tech University

3rd Year Student

Expected Dec 2026

Oshawa, ON

High School Diploma

Brooklin High School

Honor Roll

Jan 2022

Brooklin, ON

NOTABLE COMPLETED COURSES

Data Structures, Scientific Data Analytics, Business Analytics, Software Design & Analysis, Computational Science, Physics I & II, Calculus I & II.