

OTHMAN ARNAOUT

AI Solutions Architect | Agent systems, MCP tooling, workflow automation | Canadian citizen

arnaout.othman@gmail.com | linkedin.com/in/othmanarnaout | github.com/generalOtto | othman.ca

SUMMARY

AI solutions architect who designs and ships agent systems that stay in daily use: tool servers, write guardrails, human approval points, and the tests that keep them honest. Three years running a solo AI automation practice (about 40 clients in parallel, legal and tech), plus public open-source agent tooling. Earlier: Unity/C# software roles, production lead for a 30+ person team, and four years of mechanical engineering roles.

How I build: agentic engineering. I own the design, write the specs, direct AI coding agents (Claude Code), review every task and whole branch, and verify in production. Most implementation is agent-written; the decisions, guardrails and acceptance are mine.

SELECTED AGENT AND AI WORK

Brain Remote MCP server | TypeScript, MCP SDK, zod, Cloudflare Workers

Aug 2026 - Present

Public: github.com/generalOtto/claude-brain-kit (tools/brain-remote)

- Remote MCP server giving every Claude surface (web, phone, voice, coding agent) read access and guarded write access to a git-backed knowledge base. In production since August 2026, v1.0 through v1.5.1.
- Designed five tools (index, read, write, append, edit) whose descriptions teach the model the recall and write protocol. Refusals return actionable text rather than protocol errors, so the agent recovers in one retry.
- Write guardrails: path allowlist and denylist, secret-pattern scanning (including across the seam of an edit), a wipe guard, and atomic commits with compare-and-swap retry. No delete tool by design: irreversible actions stay with a human.
- 211 automated tests; every release is verified with a live production probe (write, verify, revert). A corruption bug caught in final review and a section-matching bug found in live use both became pinned regression tests.
- Made the index tool surface-aware, cutting its payload from 71k to 45k characters. Migrated hosting from a VPS container to Cloudflare Workers with the test suites mirrored case for case.

Claude Brain Kit | Bash, Claude Code plugins, skills and hooks, GitHub Actions

Jul 2026 - Present

Public, MIT: github.com/generalOtto/claude-brain-kit

- Open-source template and Claude Code plugin marketplace that gives an assistant persistent, cross-device memory from a private git repo, with no servers to run. Two-command install.
- Moved the workflow layer into on-demand skills (setup, write, to-dos, idea capture, profile) and hooks (session-start sync, a stop-time journal guard), bringing always-on context cost to about 391 tokens.
- Release gate: privacy sweep, gitleaks, a two-agent adversarial review (privacy and accuracy lenses), 65 bash tests and CI. The gate caught a secret-scanner allowlist that was suppressing stock rules and a setup script that could silently rotate a live secret.
- Retired an earlier vector-search layer (Qdrant, embeddings, a scheduled distiller) after real usage showed a curated index covered the queries. Memory writes happen in-session with the user present, not from an unattended job.

YouTube MCP | TypeScript, MCP SDK, zod, YouTube Data API v3, OAuth 2.0

Mar 2026

Public: github.com/generalOtto/youtube-mcp

- MCP server exposing 15 playlist-management tools with validated inputs and persisted OAuth tokens. Used to reorganise a 156-video playlist into five.

Quick Perplexes | Browser extension, Manifest V3, vanilla JS

Mar 2026 - Present

- Spotlight-style search overlay, live on the Chrome Web Store and Firefox Add-ons from a single build, with zero runtime dependencies.

Confidential private-client systems | details withheld

2025 - Present

- Self-hosted messaging app in production: extended an open-source federated messenger (Go server, React web, React Native iOS) with VoIP calling (WebRTC, CallKit, push) and server-enforced, fail-closed admin controls. iOS builds ship to TestFlight from an automated command-line pipeline.
- Root-caused a deployment that had been failing silently for five weeks from the build logs, and traced a calling failure to seven distinct causes.
- Scheduled media-ingestion pipeline (Playwright, systemd timers, chat alerts on every run and on failure) with a deliberate human verification step before anything is published. Production failures feed a runbook.
- On-premises inference server: GPU passthrough into a container, Ollama behind a LiteLLM gateway, access-gated through Cloudflare Tunnel.
- Designed an evaluation harness for a local-model batch translation pipeline: a structural pass gate, a repair loop, and chrF++ scoring against 72 reference translations. Smoke-tested; the full model bake-off has not been run yet.

EXPERIENCE

Digital Transformation & AI Solutions Architect | Private Establishment

Aug 2023 - Present

- Run a solo AI automation practice handling about 40 clients in parallel across legal and tech, centred on data transformation and management with AI.
- Design and implement automated systems that manage large volumes of simultaneous case projects, reducing manual effort and increasing processing speed.
- Analyse existing procedures to find bottlenecks, then deliver fitted solutions (n8n workflows, generative AI, automated reporting) and advise on tool integration so data flows end to end.
- Build and operate my own infrastructure: a self-hosted Proxmox lab and a Coolify VPS running n8n, NocoDB, MinIO and other self-hosted services behind Cloudflare.

Teaching Assistant, Capstone Production | Toronto Film School

Jan 2024 - Sep 2024

- Mentored a student team through full production of a 3D multiplayer adventure game: planning, scope, programming support and troubleshooting.

Lead Producer and Lead Programmer, Unholy Matrimony | Toronto Film School

Jun 2022 - Jun 2023

- Led 30+ people across cross-functional teams through end-to-end production of a 3D third-person action game: scope, weekly progress reviews, team structure and leads, motion capture and audio pipelines.

Software Engineer, Unity Developer | Spear Interactive (ERS Group Inc.)

Mar 2022 - Oct 2022

- Built data structures to manage large real-world datasets and generate fictional schools, teams, players and stats for a 3D sports game; created a modular UI system (C#, Unity, Plastic SCM, Jira, Agile).

Junior Game Developer | BigFatBrainStorm Group

Jul 2021 - Nov 2021

- Developed iOS and Android apps in Unity; designed a content-delivery algorithm for a children's educational game and merged several Unity projects into one platform loading asset bundles on demand from remote servers.

Scrum Master and Lead Programmer, Labtronics | Mohawk College

Jan 2021 - May 2021

- Led a four-person team and programmed a 3D electronics lab simulator for a client; ran sprint planning, backlog refinement and client milestone reviews.

Earlier engineering career | Saudi Arabia

2015 - 2020

- Doosan Power Systems (Jan 2018 - Jan 2020): Engineering Team Coordinator, then Proposal Specialist. Contributed to a winning 150 million SAR tender.
- Saudi Binladin Group (Nov 2015 - Jan 2018): Field Engineer, HVAC operations and maintenance.

EDUCATION

- **Diploma with Distinction, Video Game Design & Development**, Toronto Film School, June 2023 (President's List)
- **Graduate Certificate, Virtual Reality Multimedia Production**, Mohawk College, June 2021 (Honours, President's List)
- **BSc Mechanical Engineering**, Applied Science Private University, 2015

SKILLS

- **Agent and LLM systems:** MCP server and tool design, Claude Code plugins, skills and hooks, prompt and protocol design, write guardrails, human-in-the-loop approval design, regression tests from production failures, n8n, Ollama, LiteLLM
- **Code:** hands-on C#, C++, JavaScript, Python, HTML/CSS, SQL. Shipped through agent-directed development and review: TypeScript, Go, React, React Native, Bash
- **Platforms:** Cloudflare Workers, Tunnel and Access; Docker, Coolify, Proxmox, GitHub Actions, git
- **Languages:** English (fluent), Arabic (native)