

David Robertson

AI Agent Engineer · Backend Architecture · Context Pipelines & RAG · LLM Infrastructure

info@drobotson.pro · linkedin.com/in/david-robertson-ai · github.com/drobotson-dev · x.com/davidrobertson

SUMMARY

Full-stack AI engineer with 10+ years shipping production software and 4+ years building autonomous LLM-driven agent systems, context pipelines, and retrieval infrastructure from the ground up. Currently architecting multi-LLM orchestration with tool calling, observable agent workflows, and structured memory patterns at Waymark; previously built agent-powered microservices for private equity and developer tools. Deep hands-on experience with RAG variants, function calling, reasoning traces, and production-grade agent observability. Invited by Chris Lattner to author a technical article for Modular documenting GPU kernel work in Mojo that outperformed state-of-the-art CUDA baselines. Python-first, high-ownership builder.

CORE COMPETENCIES

Agent Systems	LangGraph · LangChain · Function calling · Tool orchestration · MCP · Reasoning traces · Agent memory · Workflow design
Retrieval & Data	RAG pipelines · pgvector · Pinecone · PostgreSQL · Heterogeneous data integration · Context engineering
Languages	Python (FastAPI) · TypeScript · OpenAI API · Anthropic API · Vertex AI · REST APIs
Infrastructure	GCP (GKE, Cloud Run) · Azure · AWS Lambda · Docker · CI/CD · Linux · Observability

PROFESSIONAL EXPERIENCE

Lead AI Software Engineer · CogniformAI · Remote

September 2024 – Present

- Architected an LLM-powered SaaS platform from the ground up (FastAPI, Postgres, GCP); own the full backend, agent infrastructure, and deployment pipeline.
- Designed a LangGraph RAG pipeline with context management and retrieval strategies that reduced developer iteration time by 50%.
- Built auto-scaling agent clusters on GKE with structured logging and observability; handled launch-day traffic with P95 latency under 400 ms.
- Concurrently take on external AI consulting engagements (see below) while maintaining delivery cadence.

Consulting Engagements (concurrent with CogniformAI)

AI Engineer · Waymark · Remote / Detroit

August 2025 – Present

- Building an autonomous agent pipeline that orchestrates multiple LLMs and generative video models, integrating heterogeneous data sources (APIs, web scraping, visual analysis, structured/unstructured data) into a single workflow.
- Designed multi-phase agent system with tool calling, context management, and structured memory: business profiling, multi-model creative generation, automated shot-list creation, and timeline assembly with inline regeneration.
- Built a natural-language AI editing assistant that parses and surgically modifies project JSON from conversational commands — function calling and reasoning traces drive re-renders without manual intervention.
- Architected complete observability layer logging all agent inputs/outputs/decisions to enable prompt evaluation, debugging reasoning traces, and systematic quality improvement.
- Working directly with Nathan Labenz (CEO); engaged as a long-term technical partner.

Generative AI Engineer · Vantager Inc. · Remote / New York

January 2025 – March 2025

- Architected and delivered a Python + LangGraph agent microservice integrating complex financial data sources for LPs managing over \$100 B in AUM; reduced report generation from weeks of manual work to under 60 seconds.
- Designed agent workflow with structured tool calling, context pipelines for heterogeneous financial data, and real-time streaming output integrated into an existing React/Next.js front-end.
- Shipped production-ready MVP in under six weeks from a standing start in an unfamiliar domain.

AI Integration Specialist & Lead Developer - FormKit Inc. - Remote

December 2023 – August 2024

- Conceived and built an autonomous agent system end-to-end (FastAPI + Docker + Postgres) that generates form schemas and UI from natural language with sub-500 ms inference latency; architecture and implementation entirely self-directed.
- Agent writes its own schemas, validates fields, and explains its decisions — reduced manual operations by 70%.
- Trained the FormKit engineering team on AI/LLM integration patterns; solution adopted as their flagship product. Public recognition from leadership.

Software Engineer - High Park Livery - Toronto

November 2022 – July 2023

- Built a custom CRM and AI-powered engagement automation suite (Vue 3/Nuxt 3); drove 225% revenue growth and 11% Google Ads CTR.
- End-to-end delivery: requirements gathering, backend payment pipeline, AI chatbot, and ongoing operations support.

AI Systems Integration Specialist - Freelance / LRS Media Group

January 2016 – September 2022

- Early adopter of LLM technology (GPT-3 beta, 2021); deployed AI solutions across tax, SEO, and logistics domains.
- Built a RAG system for an accounting firm enabling instant search across 90 years of tax bulletins; created automated LLM-powered SEO orchestration for multiple clients.

PUBLICATIONS & RECOGNITION

"How to Beat Unsloth's CUDA Kernel Using Mojo — With Zero GPU Experience"

Modular (modular.com/blog), January 2026 · Written at the invitation of Chris Lattner (Modular CEO; creator of LLVM, Clang, Swift). Documented GPU kernel optimisation work that outperformed C++/CUDA baselines across four NVIDIA GPU architectures: T4 (1.07x), L4 (1.25x), A100 (1.81x), H100 (1.60x).

"Unlocking the Power of AI in Your Vue Coding Workflow"

Vue School (vueschool.io), December 2024 · Expert guide on integrating Generative AI into developer workflows to accelerate productivity.

OPEN SOURCE

Instructor-Stream-JS — Re-engineered the streaming core of this AI ecosystem package, reducing memory usage by 40% and enabling real-time decoding for agent prompts.