Cayman Roden

Palm Springs, CA | (310) 982-0492 | caymanroden@gmail.com | github.com/ChunkyTortoise | linkedin.com/in/caymanroden

SUMMARY

AI Engineer specializing in production LLM orchestration, multi-agent systems, and RAG pipelines. Shipped a live client deployment of a multi-bot real estate sales AI (seller + buyer qualification, GHL CRM integration, Redis-backed state, deployed on Render). Built a multi-agent platform with three specialized chatbots, cross-bot handoff orchestration, and three-tier caching reducing token costs by 89%. 1,050+ automated tests across production and open-source repositories. Full-stack Python: FastAPI, PostgreSQL, Redis, Docker, GitHub Actions.

TECHNICAL SKILLS

AI / ML	LLM Orchestration (Claude, Gemini, GPT, Perplexity), RAG Pipelines, NLP/Intent Classification, Multi-Agent Systems, A/B Testing, Prompt Engineering, BM25 + Dense Hybrid Retrieval, LLM Evaluation
Languages	Python (3.9+), JavaScript/TypeScript, SQL, HTML/CSS, Bash
Frameworks	FastAPI, Streamlit, SQLAlchemy, Alembic, Pydantic, scikit-learn, XGBoost, NumPy, Pandas
Infrastructure	PostgreSQL, Redis, Docker, Docker Compose, GitHub Actions CI/CD, AWS (EC2, S3, RDS), Nginx
Tools	Git, Ruff, Pytest, Plotly, SHAP, BeautifulSoup, httpx, PyPDF2

PROJECTS

EnterpriseHub — Real Estate AI & BI Platform   github.com/ChunkyTortoise/EnterpriseHub
FastAPI + Streamlit + PostgreSQL + Redis + Multi-LLM AI | 11 CI workflows | 200+ tests

- Architected LLM orchestration layer routing across Claude, Gemini, GPT, and Perplexity with <200ms added latency and three-tier caching (memory, Redis, PostgreSQL) reducing token costs from 93K to 7.8K per workflow (89% reduction)
- Built multi-agent chatbot system with three specialized bots (lead qualification, buyer readiness, seller engagement) coordinated through a handoff service with 0.7 confidence threshold, circular prevention, rate limiting, and pattern learning
- Developed NLP intent decoding pipeline with confidence scoring, real-time CRM enrichment via GoHighLevel API, and temperature-based lead classification triggering automated workflows
- Implemented production monitoring: alerting service with 7 configurable rules and cooldowns, P50/P95/P99 latency tracker with SLA compliance, A/B testing framework with z-test significance analysis
- Built Streamlit BI dashboards with Monte Carlo simulation, sentiment analysis, and churn detection

Jorge Real Estate Bots — Production Sales AI github.com/ChunkyTortoise/jorge_real_estate_bots
FastAPI + GoHighLevel CRM + Redis + Docker | 422 tests | Live on Render — Client Deployment

- Deployed production multi-bot qualification system for a real estate investor, handling seller and buyer lead conversations end-to-end via SMS webhooks through GoHighLevel CRM

- Implemented Q1–Q4 seller qualification framework extracting property condition, price expectation, motivation, and offer acceptance; routes to Haiku classification fallback when keyword matching fails, guaranteeing conversation always advances
- Resolved critical P0 stuck-lead bugs: expanded keyword parsers across all four question handlers, added Claude Haiku fallback classifier, fixed pre-approval false-positive logic, and separated offer/timeline acceptance into independent signals
- Integrated GoHighLevel CRM — temperature tag lifecycle management (remove stale tags before applying new), custom field updates, and workflow triggers on lead qualification
- Redis + MemoryCache fallback architecture for conversation state persistence; exception handler preserves lead temperature and progress on error rather than resetting to cold

DocQA Engine — RAG Document Q&A Platform   github.com/ChunkyTortoise/docqa-engine
BM25 + Dense Hybrid Retrieval + Prompt Engineering Lab | 94 tests

- Built hybrid retrieval pipeline combining BM25 keyword search with TF-IDF dense embeddings for document Q&A, with configurable chunk strategies and relevance scoring
- Developed prompt engineering lab for systematic prompt optimization with cost tracking and response quality evaluation across multiple LLM providers
- Implemented document ingestion pipeline supporting PDF, DOCX, and plain text with automatic chunking and metadata extraction

AgentForge — Unified Async LLM Interface   github.com/ChunkyTortoise/ai-orchestrator
Claude + Gemini + OpenAI + Perplexity + Mock Provider | 27 tests

- Created unified async Python interface for multi-provider LLM orchestration with automatic failover, response normalization, and provider-agnostic benchmarking
- Built CLI tool with provider comparison benchmarks measuring latency, token usage, and cost across all supported providers

Revenue Sprint — AI Security & Optimization Suite   github.com/ChunkyTortoise/Revenue-Sprint
3 Products: Injection Tester + RAG Cost Optimizer + Agent Orchestrator | 240 tests

- Built prompt injection testing framework that identifies LLM vulnerabilities across multiple attack vectors
- Developed RAG cost optimization pipeline analyzing token usage patterns and recommending caching strategies

Insight Engine — Data Analytics Platform   github.com/ChunkyTortoise/insight-engine
Auto-Profiler + Dashboard Gen + Attribution + Predictor | 63 tests

- Built automated data profiling, marketing attribution modeling, predictive analytics (XGBoost + SHAP), and Streamlit dashboard generation from raw datasets

EXPERIENCE

AI Engineer — Independent   *January 2025 – Present*
Full-Stack AI Development | Remote

- Shipped live client deployment of a production sales AI: multi-bot qualification system (seller + buyer) integrated with GoHighLevel CRM, Redis state persistence, and Docker on Render
- Designed and built seven production-grade open-source AI repositories totaling 1,050+ automated tests with full CI/CD pipelines
- Delivered AI automation solutions for clients including chatbot systems, RAG pipelines, and BI dashboards; maintained CI green status with comprehensive test coverage across all repositories

Software Engineer — Independent Study & Projects   *January 2023 – December 2024*
AI/ML Engineering Focus | Remote

- Completed 19 professional certifications spanning AI/ML, Python, cloud computing, and software engineering (AWS, Google, IBM, DeepLearning.AI, Meta, HarvardX)
- Built Python, SQL, data structures, and machine learning foundations through structured coursework and hands-on projects, transitioning into full-stack AI development

CERTIFICATIONS

- AWS Cloud Practitioner • Google IT Automation with Python • IBM AI Engineering • Meta Back-End Developer
- DeepLearning.AI: ML Specialization, TensorFlow Developer • IBM: Data Science, Full Stack Development
- Google Data Analytics • Google Advanced Data Analytics • HarvardX CS50 Python