

I deliver regulated credit-risk data inside a bank by day and run an agent platform by night: planner, workers, and verifier as separate stages, model routing and prompt caching by task, and no job called done without evidence. Language models in production since 2022. Three companies founded. Teams of seven and six led.

≈26M

records a day through production credit-risk pipelines

19 yrs

of history corrected in place, 2006 to 2025

1,002

scheduled agent runs in 18 hours, each with evidence

40,000

leads at \$0.29 each from an autonomous creative loop

\$100K+

raised in cash and prizes for ReCap Budget

01 EXPERIENCE

USAA · Data Engineer III · Data Engineering Intern, summer 2023 June 2024 - present
Snowflake · dbt · Python · SQL · GitLab CI/CD · Control-M

- Delivered three major consumer-lending features in the second half of 2025, including a legacy pipeline retirement and the FICO 10T data asset for CECL, both on the regulatory deadline with reporting uninterrupted.
- Corrected and backfilled nineteen years of credit-risk history in place with targeted comparisons instead of rebuilds; built the reusable validation query the team and its validators now run.
- Sit on the Data Science and Engineering Guild that decides how generative AI is adopted across the bank. Mentor and onboard engineers; wrote dozens of SKILL.md guides for testing, Snowflake debugging, and scheduling.
- As an intern, took a credit-card underwriting feature from requirements through the production control gate and built an offline RAG assistant (ChromaDB, Falcon-40B) presented to hundreds of employees.

Guadalupe Automation · Founder and CEO January 2026 - present
Python · FastAPI · Hermes · Postgres + pgvector · Docker · Windmill · AWS

- Store-close reconciliation for a multi-store retailer: fifteen deterministic parsers and Python reconciliation; a model lane reads only scanned evidence against closed schemas, every proposal goes to human review, exports need role-gated approval. June close: 295 of 315 checks reconciled, zero open amount differences, 857 tests.
- SecondHome, distressed-property lead qualification: twenty county sources normalized with the source kept on every field, seventeen bounded agent skills, a versioned deterministic score, and a written call rationale on every lead. 1,039 tests.
- Operate isolated Hermes agent environments for six clients and six private users on per-client VPS and Docker stacks. Own technical delivery end to end with one employee and a sales partner.

Fit & Flourish · Co-founder, marketing and automation lead 2024 - 2025
Meta Ads API · n8n · OpenAI · Python

- Built an autonomous Meta Ads creative loop: generate variants, multimodal brand-and-policy check, human approval in Discord, deploy, kill underperformers daily. 40,000 waitlist leads; cost per lead from \$2.19 to \$0.29. Led a team of six.

ReCap Budget · CEO and ML Engineer December 2021 - September 2023
AWS Lambda · S3 · Textract · FastAPI · GPT-3 · PyTorch

- Led seven developers and raised \$46,000 cash plus about \$60,000 in-kind prizes for a serverless personal-finance app on 25+ AWS services.
- Built the event-driven receipt pipeline (S3 to Lambda to Textract behind a CNN gate) and fine-tuned a GPT-3 classifier on 40,000 curated items, 100+ categories at over 85% reported accuracy. Shipped schema-shaped JSON from LLMs before function calling existed; bias-mitigation work on the fine-tuned advice model.

“You bring something we didn’t have before.” USAA manager, 2026

04 SKILLS

Agents and models. Hermes, LangChain, LangGraph, pgvector, ChromaDB, Pinecone, Redis, prompt caching, model routing, Weights & Biases, Weave, fine-tuning (GPT-3, Qwen3, PyTorch, Unsloth), Claude Code, Codex, Cursor, Gemini CLI, Bedrock.

Data and platform. Python, SQL, FastAPI, Snowflake, dbt, Postgres, DuckDB, AWS Lambda, S3, Textract, Docker, GitLab CI/CD, Control-M, n8n.

02 AGENT PLATFORM

Joy 2025 - present
My agent, operating daily. Hermes runs the loop; GBrain holds cited memory on pgvector; a router fine-tuned on Qwen3 8B sends each job to a cheap, frontier, local, or long-context model through a caching gateway; a staged overnight pipeline plans, dispatches, and verifies, and refuses “done” without evidence. 1,002 scheduled runs in 18 hours, 279 reusable skills, one of thirteen agents on infrastructure I run.

Cyrus-v2 2026
Agent orchestration. Five coding agents (Claude Code, Codex, Gemini CLI, OpenCode) behind one relay with per-model token and cost metering, compute-pool budgets, a scheduler with cooldowns and circuit breakers, and a lifecycle audit before any work item closes. A PM agent grooms the backlog between dispatches.

Earlier 2023 - 2026
Cyrus on the first LangChain release; TinyVDB, an in-process vector store; Splice, contract-bound LLM code execution; Cortex, always-on voice to structured memory; VibeTest, deterministic persona evaluation.

03 HOW I BUILD

Deterministic boundaries. Exact math, schemas, permissions, and acceptance checks live in code. Models get extraction, synthesis, classification, and bounded planning.

Planner, workers, verifier. Separate stages with separate authority. Every run leaves a ledger entry. Retries, fallback chains, leases, and fail-closed defaults.

Evaluation. Weights & Biases for models, Weave for agents. Bias-mitigation on a fine-tuned advice model. Provenance and confidence on every scored record.

Humans own the irreversible. Outreach, spending, deletion, and approvals stay with a person. A false positive beats a false negative when money is involved.

05 EDUCATION

B.S. in Computer Science, Trinity University, San Antonio, 2024.