

Timothy Waldin

timothy@waldin.net | (415) 676-9427 | linkedin.com/in/twaldin | github.com/twaldin | tim.waldin.net

EXPERIENCE

Software Engineer | Lindy.ai — Agents team

May 2026 – Present

- Shipped Online Scoring: evaluation system reading and grading millions of production agent traces per week with LLM judges orchestrated by Temporal workflows
- Implemented GEPA prompt optimization in production systems handling thousands of user agents

PROJECTS

Agentic Engineering Suite *TypeScript, Python, Bun, tmux, GEPA* github.com/twaldin

2025 – Present

- **tmux-orchestrator + claudecord**: Claude Code plugins spawning multi-agent teams in isolated tmux windows with git-worktree isolation, send-keys messaging, and Discord routing so the fleet is reachable from a phone — the genesis prototype
- **flt**: generalized the orchestrator to 6 coding CLIs (Claude Code, Codex, Gemini CLI, Aider, OpenCode, SWE-agent); raw-ANSI TUI with damage-tracked screen buffer, vim keybinds, per-CLI spinner detection, auto-approved permission prompts, inter-agent inbox. Published as [@twaldin/flt-cli](https://github.com/twaldin/flt-cli) on npm
- **harness**: extracted the multi-CLI adapter layer into a published Python (`harness-cli`) and TypeScript (`@twaldin/harness-ts`) library covering 13 coding CLIs with byte-level fixture parity between the two implementations
- **hone**: GEPA-based prompt optimizer that drives coding-CLI subscriptions through harness as its mutator instead of paid API keys; lifted Claude Haiku 4.5 from 65% to 85% on a 9-bug GitHub hold-out (+20pp, no regressions) for ~\$1 in mutator tokens
- **AgentElo**: pairwise Bradley-Terry leaderboard ranking 148 agents across 6 harnesses on real GitHub PRs — ~1B tokens, \$642 spend, 3.5K verified runs. Read-only baseline at tim.waldin.net/agentelo CLI now runs locally to rank your agent against the bundled snapshot (public submissions closed after Stanford's Terminal-Bench 2.0 + Harbor shipped equivalent hosted infra)
- **harness-bench**: current focus — benchmark the **scaffold** not the model by holding the model fixed and swapping the harness across SWE-bench Live and Terminal-Bench 2.0

CS2 Trade-Up Bot *TypeScript, React, Express, PostgreSQL, Redis* tradeupbot.app

2025 – Present

- Full-stack market arbitrage platform analyzing profitable CS2 trade-up contracts across 3 marketplaces in real time with 100+ daily active users
- Discovery engine evaluating 100,000+ listing combinations per cycle with Steam OpenID auth, Stripe subscriptions, Cloudflare Turnstile, real-time WebSocket feeds, 17-currency display

term-site *Next.js, Docker, Socket.IO, xterm.js* tim.waldin.net

2025

- Terminal-based portfolio where visitors explore projects inside isolated Docker containers through a browser shell; URL paths (`/blog/<slug>`, `/projects/<name>`) deep-link to shell commands

StudySpot *Next.js, Cloudflare Workers, pgvector, Claude API* studyspot.us

2025

- Co-founded AI study platform using RAG + Claude to answer questions from uploaded course materials; pgvector semantic search, streaming responses, auto-generated flashcards and quizzes

TECHNICAL SKILLS

Languages: TypeScript, JavaScript, Python, C, C++, Java

Frameworks: React, Next.js, Node.js, Express, Tailwind, Socket.IO, shadcn/ui

Tools: Git, Docker, PostgreSQL, SQLite, Redis, nginx, Cloudflare Workers, Vercel, Bun, uv

AI: GEPA, DSPy, ACE, RLM; Bradley-Terry / Glicko-2 ranking; SWE-bench Live, Terminal-Bench 2.0, HumanEval; agent benchmarking and prompt optimization

EDUCATION

Purdue University | Polytechnic Institute

On leave since May 2026

B.S. Web Programming and Design (on leave of absence)

Relevant Coursework: C Programming (CS 159), Microcontroller Programming in C (ECE 264), Intro to Web Development (CGT 141)

Leadership: Founded high school robotics team; competed at FIRST Robotics Competition (FRC) World Championships (2×)