

KEITH SOYKA

AI Systems Builder | Implementation Advisor | Technical Mentor

keithsoyka@gmail.com · KeithSoyka@GestaltView.onmicrosoft.com · 646-513-6410 · New York, NY
linkedin.com/in/keithsoyka413 · github.com/digitalconsciousness

PROFESSIONAL SUMMARY

Hands-on AI systems builder and technical mentor with 16+ months of intensive production LLM work and 22 years of frontline operations leadership, training, and coaching. Designed and shipped GestaltView, a production AI continuity platform spanning Python and TypeScript, REST APIs and serverless functions, large language models and generative AI tooling, prompt engineering and systematic model evaluation, retrieval-augmented generation (RAG) with 768-dimensional embeddings and pgvector vector databases, and cost-aware multi-model orchestration across OpenAI, Anthropic Claude, Google Gemini, Groq, and Hugging Face. Daily practice includes AI-assisted engineering workflows, multi-model output comparison, data annotation and evaluation rubric design, and governance standards separating verified source from model inference. Recognized as a quarterfinalist (top ~4% of 2,300) in the 2025 Pepperdine Most Fundable Companies competition.

CORE SKILLS

- **LLMs & Generative AI:** OpenAI GPT, Anthropic Claude, Google Gemini, Groq, Hugging Face; multi-model routing and cost-aware orchestration; AI-assisted engineering workflows
- **RAG & Retrieval:** retrieval-augmented generation, 768-dim embeddings, pgvector vector databases, hybrid semantic + keyword search, chunking strategies, bounded retrieval
- **Prompt Engineering & Evaluation:** systematic prompt design, multi-model output comparison, model evaluation rubrics, data annotation, failure-mode analysis, RLHF concepts
- **Languages & Frameworks:** Python (NumPy, Pandas; PyTorch/TensorFlow concepts), TypeScript, SQL/Postgres, REST APIs, serverless (Vercel), RAG patterns equivalent to LangChain/LlamaIndex workflows
- **AI Systems:** agent orchestration, persona/embodiment systems, voice-ingestion pipelines, artifact generation and rendering pipelines, session memory and continuity architectures
- **Mentoring & Training:** 22 years onboarding, training, cross-training, and coaching operational teams up to 75 people; curriculum-style technical explanation; 1:1 and group facilitation
- **Data & Governance:** evidence standards (verified / partial / planned / unsupported), provenance tracking, consent-aware storage, private-by-default design, RLS security policies

EXPERIENCE

Founder & Product / AI Systems Lead | GestaltView — May 2025 – Present · New York, NY

- Solo-designed and shipped a production AI platform (Next.js/Supabase/Vercel, Stripe billing): 722 TypeScript files, 217 Postgres tables, 299 RLS policies, 75 database functions.
- Built a hybrid retrieval engine: one query fans into five parallel streams (semantic + keyword over knowledge and skill corpora plus memory), fused with RRF ranking and graceful degradation.
- Implemented cost-aware LLM cascade (local/free-first routing across 7 providers) with per-response provider and cost reporting; Braintrust-traced observability.
- Created a schema-driven persona compiler (25 embodiment profiles) with drift detection and voice-preservation scoring; pre-LLM intent/emotion/flow analysis on every request.
- Diagnosed and repaired a 12-fracture artifact render pipeline across client, API, database, and storage boundaries; documented repairs with tests (evidence-first repair packages).
- Built voice-ingestion pipeline (audio → transcription → themes → linked memory), generative engine (16 artifact types), and orchestration layer routing content by kind across product surfaces.

Operations Leadership | Multi-site retail operations — 2002 – 2024

- 22+ years frontline P&L ownership: labor planning, inventory, multi-site management; trained, onboarded, and coached teams of up to 75.

- Career-long practice translating complex systems into teachable procedures — the same skill applied now to AI mentorship.

RECOGNITION

- Pepperdine Most Fundable Companies 2025 — Quarterfinalist, top 100 of 2,300 startups (top ~4%).

PORTFOLIO & LINKS

- **GitHub:** github.com/digitalconsciousness — production repositories
- **LinkedIn:** [linkedin.com/in/keithsoyka413](https://www.linkedin.com/in/keithsoyka413)

Available for advisory, consulting, and bespoke implementation alongside the venture; structured so building stays first.