

Ken Faiman

AI Systems Engineer | Agent Orchestration | Eval Harnesses | AI Pipeline Tooling

Fullerton, CA | kfaiman@gmail.com | linkedin.com/in/kfaiman | github.com/sushiHex | US Citizen
faiman.com | Open source: [mainframe-mcp](#), [claude-oracle](#), [hermes-realtime](#) | [IMDb](#)

PROFILE

AI systems engineer specializing in agent orchestration and eval harnessing: the self-improving feedback loops that make agent systems reliable. At Meta Reality Labs, built the agent tooling an art organization ran on, from an agent-swarm research system adopted by 50+ users to genAI asset CI across 1M+ assets, and trained the team to use it.

Before that, two decades building production pipelines, tools and simulation for film and games (*The Matrix*, DreamWorks, Blizzard). Ships production systems in Python and C++17, test-driven and in CI.

SELECTED HIGHLIGHTS

- ▶ **Agent systems with evals.** Agent-swarm research system for 50+ users, hardened by a 15-axis self-improving eval harness.
- ▶ **Agent-first tooling.** Designs CLIs with the LLM as a first-class user: streaming JSONL, `--help` as spec, safe dry runs.
- ▶ **AI pipeline CI.** C++ FBX-to-USD library with validation and auto-fixing, run as an agent-ready CLI on 1M+ assets.
- ▶ **Open source.** An MCP server with GPU RAG and a 50-query eval gate; a research orchestrator 3.2x cheaper than Opus-only.
- ▶ **ML safety in production.** Built a render system that ran ML image-safety checks across 200,000+ 2D and 3D assets.

CORE SKILLS

Agents & Orchestration

- Multi-agent orchestration
- MCP servers (authored)
- Claude Code, Agent SDK, Codex
- Tool use & autonomous loops
- Agent-first CLI design

Evals & Retrieval

- Eval harnesses, LLM-as-judge
- Multi-metric benchmarking
- RAG, hybrid BM25 + vector
- Reranking & NLI grounding
- Regression gates & observability

Engineering & ML

- Python, C++17, TypeScript
- Test-driven development, CI
- PyTorch, diffusers, PEFT
- LoRA fine-tuning, LightGBM
- USD / OpenUSD pipelines

EXPERIENCE

Senior Technical Artist, GenAI & Avatars | Meta Reality Labs

2022 - 2026 | Meta Avatars and Horizon | Rated "Exceeded Expectations" (2025)

- Built an agent-swarm research system, taken from prototype to company-shared and adopted by 50+ users, hardened through a self-improving eval harness (15-axis rubric, 200+ full-round iterations).
- Developed custom agent frameworks on Claude Code and trained the team's Technical Artists to adopt agentic workflows.
- Wrote an overnight agentic task looper that built the team's full-stack React asset-review web app end to end across 500+ headless tasks, debugging from its own error logs and deciding whether to retry or skip each stalled task.
- Engineered a thread-optimized C++ library converting FBX to engine-compliant OpenUSD with validation and automatic fixing, shipped as a CLI and a CI-pipeline node across 1M+ vendor and genAI assets.
- Designed that CLI with the LLM as a first-class user: JSONL output, `--help` as the agent's spec, a dry-run mode, and arguments grown from agent trial runs.
- Developed a headless OpenUSD render system (10k assets per batch) and ran ML image-safety checks across 200,000+ assets.
- Created a Claude-based system that auto-validated and conformed 100 GenAI fantastical-character rigs as training data for a physics-based ML rigging model.
- Landed 28 pipeline-node changes with an AI coding assistant, turning weeks of scheduled Engineering into hours.

- Sped up server-side asset I/O at least 2x by parallelizing it and benchmarking the storage APIs.
- Fixed a production hair-LOD failure on 4,664 user requests, cut to zero, and reported it at the Engineering incident review.
- Helped build the next-generation Meta Avatars: real-time rigging, deformation and node/graph "blueprint" systems.

Senior Rigging & Simulation Artist II (Lead) | [Blizzard Entertainment, Cinematics](#)

2015 - 2022 | Cinematics for all seven Blizzard game franchises

- Rig/sim show lead for 10+ riggers on Overwatch, World of Warcraft and Diablo cinematics.
- Designed the Nucleus-based sim framework and PyQt tools for Stitches, the studio's modular Python rig-build system.

Character FX Lead / Senior Simulation Developer | [DreamWorks Animation](#)

2005 - 2015 | Eight feature films, including *How to Train Your Dragon* and *Kung Fu Panda*

- Show lead for 25+ Character FX artists; pioneered a cloth-based skin solver for self-collision and contact.
- Worked with R&D on two physics solvers, one later presented at SIGGRAPH; wrote core Python tools for the department.

Crowd / Motion Capture Technical Director | [Rhythm & Hues Studios](#)

2004 - 2005 | *The Chronicles of Narnia: The Lion, the Witch and the Wardrobe*

- Procedural retargeting for the film's autonomous crowd agents; wrote the team's core artist tooling in Tcl/Tk.

Facial Capture / FX Technical Director | [ESC Entertainment](#)

2002 - 2004 | *The Matrix Reloaded*, *The Matrix Revolutions*

- Core contributor to Universal Capture (VES Award 2004; Academy Sci-Tech Award 2015; team awards); co-author of "Making of The Superpunch", ACM SIGGRAPH 2004 Sketches.

Earlier: 3D Game Artist, Activision (2000)

OPEN SOURCE & PERSONAL R&D

- **mainframe-mcp**: MCP server with a GPU-backed RAG engine (embeddings, cross-encoder reranking, NLI grounding, hybrid BM25/vector) and a 50-query eval harness with a regression gate.
- **claude-oracle**: pip-installable research orchestrator; 80 parallel Haiku scouts in eight isolated chains, 3.2x cheaper per research dimension than eight Opus agents.
- **hermes-realtime**: realtime conversation runtime on LiveKit that separates latency-critical speech from long-running autonomous work; 937 tests, strict mypy.
- **Adventure Game Control**: agentic layer over clean-room C++17 game interpreters, with local-to-cloud confidence routing, RAG grounding, an embedded MCP server and a multi-provider voice stack; 3,800+ tests in CI.
- **Diffusion fine-tuning**: LoRA on FLUX generating full typefaces from two seed glyphs, scored by a TrOCR + LPIPS + DINOv2 eval harness.
- **Applied ML**: 36 LightGBM quantile models over 1.4M game logs with a 369-feature pipeline and a MILP optimizer; 89.9% backtest win rate.
- **MotionTerminal** (Unreal Engine 5.8): four rigs on one Animation Blueprint Template, with a C++ plugin that diffs every AnimGraph per commit; authored agentially over MCP.

EDUCATION

UCLA, B.A. Film, Television & Animation, Honors (2000) | [Gnomon School of Visual Effects](#), Maya