

Ken Faiman

Developer Advocate & Applied AI / Agentic Systems Engineer

Fullerton, CA (greater LA / Orange County) · kfaiman@gmail.com · (714) 686-0465 · US Citizen (no sponsorship needed)
linkedin.com/in/kfaiman · GitHub: sushiHex (open source: [mainframe-mcp](#), [claude-oracle](#)) · IMDb: nm1363419

SUMMARY

Applied-AI engineer specializing in **agent orchestration and iterative-improvement harnessing** - the evaluation harnesses and self-improving feedback loops that make agent and model systems reliable. At Meta Reality Labs I built custom agent frameworks on Claude Code, MCP servers with RAG, multi-agent orchestration, and the genAI asset CI that gated them, then trained the team's technical artists and drove adoption across a 50+ user organization. I author my own open-source infrastructure (an MCP server with GPU RAG, and a multi-tier research orchestrator), build multi-metric eval harnesses (LLM-as-judge + model metrics), and work in real production code: modern C++17 and Python, test-driven, in CI. The combination is rare: I ship the agent systems and the harnesses that prove they work.

CORE COMPETENCIES

Agentic AI orchestration (multi-agent, autonomous agents) · AI agent frameworks & protocols (MCP, claude-agent-sdk, LangChain-class patterns) · LLM tool-use · RAG / hybrid retrieval · **eval harnesses (multi-metric + LLM-as-judge), iterative-improvement & feedback loops, observability** · Hugging Face ecosystem (diffusers / PEFT / TrOCR / DINOv2) · prompt engineering · developer enablement, training & community adoption · technical writing & conference speaking · Python & modern C++17 · framework authorship (MCP server, multi-agent orchestrator)

DEVELOPER ADVOCACY & ENABLEMENT

- **Open source + authored frameworks:** Author of [mainframe-mcp](#) (an MCP server with a GPU-backed RAG engine) and [claude-oracle](#) (a pip-installable multi-tier research orchestrator: parallel Haiku scouts, Sonnet synthesis, Opus architect). Both are public (MIT): mainframe-mcp ships a 50-query eval harness and regression gate in-repo; claude-oracle delivers roughly 10x research breadth at a fraction of Opus-only cost.
- **Developer enablement at Meta:** Trained the team's technical artists to adopt agentic AI workflows; built a swarm-based multi-agent research system adopted across the team (50+ users); built a Claude-based shared workspace that aggregated a 5-person pod's work across repos into a single queryable context. Drove tooling adoption org-wide, not just authored it.
- **Technical speaking & writing:** SIGGRAPH 2004 conference talk, "Making of The Superpunch" (Universal Capture facial system, *The Matrix*). Actively building a public developer presence around agentic AI: authored frameworks, a documented portfolio, and technical write-ups.

SELECTED AI / ML ENGINEERING

- **MCP server + GPU RAG (authored):** Authored an MCP server with a GPU-backed RAG engine on a single RTX 3090 - embedding model + cross-encoder reranker + NLI grounding + hybrid BM25/vector retrieval - for agent tool-use and grounded answers.
- **Multi-agent orchestration & autonomous infrastructure (Meta):** Built a multi-agent research system adopted across the team (50+ users), hardened through a self-improving eval harness (15-question scoring matrix, 200+ full-round iterations); an overnight remote-server task looper running 500+ headless coding/build tasks; and a multi-session orchestrator coordinating 5+ concurrent Claude sessions across remote servers.
- **Agentic frameworks & CI tooling (Meta):** Built custom agent frameworks (Claude Code) for AI-assisted research and coding, plus a thread-optimized C++17 library converting FBX to engine-compliant USD with full validation and automatic non-compliance fixing - shipped as an LLM-integrated CLI and a CI-pipeline node across 100,000s of genAI and vendor assets; plus a Claude-based system that auto-validated and conformed character assets for ~100 characters used to train a physics-based ML rigging model. Plus a fully multi-threaded, CPU-based HdStorm + Vulkan (OpenUSD) system doing direct PBR renders of 10k-asset batches on devservers.
- **Agentic systems R&D - retro (personal, private repo, walkthrough on request):** Agentic layer over clean-room C++17 game interpreters - local-SLM→cloud confidence routing, anti-hallucination RAG grounding, multi-

agent autopilot, an embedded MCP server, and a multi-provider voice stack (STT/TTS, provider routing, per-user cost metering). 737 commits, 3,800+ tests, CI.

- **Diffusion fine-tuning on the Hugging Face stack (personal R&D):** LoRA fine-tuning of FLUX (PyTorch / diffusers / PEFT) generating full typefaces from two seed glyphs; custom multi-metric eval harness (TrOCR + LPIPS + DINOv2) benchmarking quality, latency, and cost; Gradio demo.
- **Applied ML at scale (personal R&D):** Quantitative ML engine - 36 LightGBM quantile models over 1.4M game logs with a 369-feature pipeline, knowledge distillation, and a MILP optimizer; 222 accuracy-gated tests, 89.9% backtest win rate.

EXPERIENCE

Meta - Reality Labs · Senior Technical Artist (IC5) · 2022–2026

Built the AI and agent tooling the art organization ran on, and led its adoption; plus real-time avatar systems for Meta Avatars and Horizon.

- Built the agent frameworks, MCP/RAG tooling, and multi-agent systems above (Claude Code), and trained the team's technical artists plus a 50+ user org to adopt agentic AI workflows.
- Built and maintained real-time avatar systems - node/graph ("blueprint") backend logic and rig automation in a production game engine - the production domain that grounds the AI tooling.
- Rated "Exceeded Expectations" in the most recent review for technical contributions and leadership; role eliminated in a structural May 2026 reduction (non-performance).

Blizzard Entertainment - Cinematics · Senior Rigging & Simulation Artist II (Lead) · 2015–2022

- Architected modular, graph-based build components (Python rig modules) deployed across all cinematic pipelines (five franchises); built the simulation methodology, tools, and pipeline used studio-wide.

DreamWorks Animation · Character Effects Artist / Senior Cloth Developer · 2005–2015

- Hero-character cloth, hair, and skin across eight feature films (How to Train Your Dragon, Kung Fu Panda, Puss in Boots, and more); pioneered cloth-based skin solvers.

ESC Entertainment · Facial Capture / FX Technical Director · 2002–2004

- Built facial rigging and capture on the award-winning Universal Capture system for *The Matrix Reloaded* and *The Matrix Revolutions* (Academy Sci-Tech / VES team recognition). SIGGRAPH 2004 talk: "Making of The Superpunch."

Earlier · Rhythm & Hues (Motion Capture TD, 2004–2005) · Activision (3D Game Artist, 2000)

SKILLS

- **AI agents & frameworks:** Agentic frameworks (Claude Code, claude-agent-sdk, OpenClaw) & custom agent tooling · multi-agent orchestration · MCP servers (authored) · LangChain-class agent/tool-use patterns · RAG / hybrid retrieval (BM25 + vector, cross-encoder reranking) · LLM-as-judge & eval/observability
- **ML / GenAI:** Hugging Face ecosystem (diffusers, PEFT, transformers, TrOCR, DINOv2) · LoRA fine-tuning (FLUX) · LightGBM / quantile models · embeddings · GPU model serving (RTX 3090) · genAI asset CI
- **Code:** Python · modern C++17 (test-driven, CI) · JavaScript/TypeScript · C# · OpenMaya API · MEL
- **Infra / Tools:** Git / CI · Linux · VS Code · CLIs & SDKs · Gradio · USD/OpenUSD · Maya · Houdini
- **Advocacy / enablement:** Technical writing & conference speaking (SIGGRAPH) · developer training & documentation · cross-team adoption of new tooling · framework authorship (MCP server, multi-agent orchestrator)

SELECTED CREDITS & EDUCATION

- **Open source:** mainframe-mcp (MCP server + GPU RAG, MIT, github.com/sushiHex/mainframe-mcp) · claude-oracle (multi-tier research orchestrator, MIT, github.com/sushiHex/claude-oracle)
- **VR / real-time:** avatar/character + AI tooling for Meta Avatars and Horizon at Meta Reality Labs (2022–2026)
- **Games cinematics (Blizzard):** World of Warcraft, Overwatch, Diablo, Heroes of the Storm, Hearthstone (2015–2022)
- **Feature film / animation:** The Matrix Reloaded / Revolutions (2003) · DreamWorks (Kung Fu Panda, How to Train Your Dragon, Puss in Boots, and more, 2006–2015)
- **Education:** UCLA — B.A., Film, Television & Animation · Gnomon School of Visual Effects — Maya