Zhe Zhang

Moraga, CA (SF Bay Area) · +1 (510) 518-0963 · wenatuhs@gmail.com · linkedin.com/in/wenatuhs · github.com/zhe-slac · wenatuhs · zhezhang.info

SUMMARY

Staff Engineer at SLAC with 7 years shipping optimization and ML systems that run live on particle accelerators (physics PhD). Created **Badger**, the open-source ML optimizer used in control rooms at 7+ facilities and in LCLS-II commissioning; now building Claude Code/MCP agents that tune accelerators under operator approval — validated on live LCLS shifts — and running LLM-agent autoresearch on my own Bayesian optimizers. First author of 10+ publications.

EXPERIENCE

SLAC National Accelerator Laboratory · Menlo Park, CA 2019 – Present

Staff Engineer (2023–) · Data Engineer, ML Initiative (2021–23) · Research Associate (2019–21)

- **Created Badger** (xopt-org/Badger), the open-source ML optimization platform for accelerator control rooms, and led it through 2025: first author (IPAC'22), ~54% of 1.9k commits, 19 releases, led its LCLS beamline testing. Adopted at 7+ facilities on three continents (SLAC, DESY, BNL, ESRF, Argonne, Fermilab, KIT), used in LCLS-II commissioning; reaches optimal beam lifetime at ESRF 7× faster than conventional scans (>50× vs a human operator). (2021–)
- **Built Otter**, a Claude Code + MCP agent that tunes accelerators through Badger under LBNL's Osprey safety harness — 35 MCP tools (~5.4k LOC), hook-enforced approvals, offline knowledge packs — and extended Badger with its agent-facing control surface (routine hot-swap, live evaluation stream, external stop). (2025–)
- **Validated Otter on live machines**: on LCLS shifts (Jun 2026) it drafted 44 operator-approved tuning routines (~3 min from ask to beam) and ran a hands-off campaign taking the FEL from a 0.47 mJ sag to the evening's 1.82 mJ peak (above the manual baseline) in 88 evaluations / ~19 min, zero constraint violations; on FACET-II it completed five 6-phase injector bring-up campaigns end to end with autonomous retry.
- **Ran LLM-agent autoresearch on DAMA-BAX** — first external users of LBNL's framework: Claude agents autonomously rewrote the optimizer over hundreds of experiments; on a benchmark hardened against reward hacking, they rebuilt a deliberately crippled optimizer back to the hand-tuned reference in a single experiment and surpassed it within seven (hypervolume 11.40 vs 10.80).
- **#2 contributor** (249 commits) to **Osprey**, LBNL's open-source agentic-control framework (67 stars): shipped its Textual TUI (~5.5k LOC), unified typed-event / LLM token-streaming layer and provider adapters.
- **First-authored DAMA-BAX** (arXiv:2511.17850): applied multipoint Bayesian Algorithm Execution — via my open-source mpBAX framework — to storage-ring aperture design, matching NSGA-II Pareto fronts with 40–90× fewer simulations end to end (>400× in the optimization loop; ~100 vs ~38k CPU core-hours).
- **Built a plug-and-play RAG chatbot for research groups** — a widget any group embeds on its site to answer questions from its own papers and docs: FastAPI/SSE streaming, multi-knowledge-base embedding retrieval (OpenSearch/Qdrant), 4-provider LLM routing, always-on RAGAS evaluation; in production on SLAC's Kubernetes cluster (open-source OuRAGboros). (2025–26)
- **Built ReconNN**, a PyTorch neural surrogate (GNN / sparse 3D CNN) for neutrino event reconstruction, ~300× faster than per-event gradient-descent inference (A100), trained self-supervised through a differentiable simulator; ran 8k-job Slurm Monte-Carlo production. (2024)
- **Sole-built Teeport**, a WebSocket platform for remote online optimization (SLAC, Argonne; Frontiers in Big Data 2021); co-developed **MG-GPO**, a GP-guided evolutionary optimizer that beat PSO on the live SPEAR3 ring (MLST 2020); first-authored **RCDS-S**, a safe optimizer tracking drifting optima under safety thresholds (PRAB 2022); 2nd author of **CoAD**, label-free anomaly detection (MLST 2024).

Visual3D Inc. · Beijing 2017 – 2019

Algorithm Engineer (2018–19) · Full-stack Engineer (2017–18)

- Core developer of a mixed-reality surgical-navigation robot system (registration, multi-client scene sync); trained U-Net-family segmentation models reaching Dice 0.92 on clinical cases.

SELECTED PROJECTS

quantecho (2026–) — Solo-built, self-hosted full-stack platform (~84k LOC: Next.js 15 / React 19 / Socket.IO + FastAPI sidecars) with a stock backtest lab (in-browser Python editor, S&P 500 + CSI 300/500 bars, 90 tests), developed spec-first with Claude Code.

SKILLS

ML & Optimization Bayesian optimization (GP/BoTorch via Xopt; EI/UCB/TuRBO/MOBO), Bayesian Algorithm Execution (BAX), evolutionary & safe online optimization, neural surrogates, GNN / sparse CNN, anomaly detection, PyTorch, GPU training (A100/V100)

LLM & Agents Claude Code / Agent SDK, MCP servers & tool use, agent safety harnesses (hooks/approvals), RAG & embeddings (LangChain, OpenSearch/Qdrant), LLM evals (RAGAS, LLM-as-judge), OpenAI/Anthropic/Gemini APIs

Systems Python (12+ yrs), TypeScript/Node (5+ yrs), C++/C#, FastAPI, asyncio, WebSocket/SSE, distributed systems, Docker, Kubernetes, Slurm/HPC, EPICS, CI/CD; React/Next.js, Vue, PyQt

EDUCATION

Ph.D., Accelerator Physics, Tsinghua University, 2017 (visiting: UCLA 2014, PSI 2013) · **B.Sc., Engineering Physics**, Tsinghua University, 2011 (rank 5/150)

SELECTED PUBLICATIONS

Z. Zhang et al., Multipoint BAX for lattice design, *arXiv:2511.17850* (2025) · **Z. Zhang** et al., Badger: the missing optimizer in ACR, *IPAC'22* · **Z. Zhang**, M. Song, X. Huang, RCDS-S drift compensation, *PRAB* 25, 122801 (2022) · **Z. Zhang**, X. Huang, M. Song, Teeport, *Front. Big Data* (2021) · **Z. Zhang***, M. Song*, X. Huang, MG-GPO online optimization, *MLST* (2020; *equal) · *full list: Google Scholar*