

MATTHEW PENG

Torrance, CA (U.S. Citizen) | (626) 840-3897 | mattpeng@gmail.com | <https://www.linkedin.com/in/mzpeng/>

SKILLS

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

Data/ML: LangChain, LangGraph, OpenAI/Whisper, Ollama, Pinecone, Pandas, NumPy

Infrastructure: MySQL, AWS S3, Redis, NGINX, Docker, GitLab CI, REST APIs

Frameworks: Plotly Dash, Next.js, React, Spring Boot

WORK EXPERIENCE

Medtronic (MiniMed): Software Engineer II

07/2023 – Present

- Solely own and operate an end-to-end production platform used by 107 researchers across ~8 R&D teams, processing ~29M records per active month across automated ingestion, MySQL storage, backend computation, and web serving.
- Scaled production serving after a 3× increase in data volume, cutting worst-case page-load latency ~50% and increasing concurrent calculation capacity 5× by designing NGINX load balancing across 5 workers, server-side caching, vectorization, early filtering, and elimination of redundant computation.
- Improved production reliability through fault isolation, automated validation, and actionable error handling, reducing manual intervention ~85% and preventing malformed inputs from blocking independent workloads.
- Refactored shared Python infrastructure used across 5 R&D teams, replacing duplicated business logic and hard-coded assumptions with standardized base classes, reusable modules, and configurable interfaces across ~50 calculations.
- Reduced feature development and validation from ~3 weeks to <1 week and developer onboarding from ~1 month to 2–3 days by standardizing interfaces, improving documentation, and adding GitLab CI regression tests against known expected and legacy outputs.
- Modernized legacy MATLAB infrastructure into Python, preserving output correctness against existing implementations while improving speed, maintainability, testability, and developer accessibility.
- Adapted an open-source GPT-based GitLab merge-request reviewer with repository context, team-specific review criteria, and automated MR comments to prototype AI-assisted detection of bugs, performance issues, missing tests, and code-quality regressions.

Medtronic (MiniMed): Software Engineer Intern

06/2022 – 08/2022

- Prototyped the Python analytics workflow that became the foundation for later production sensor-analysis systems, including automated metric generation and AWS S3-backed custom calculations

National Security Innovation Network: Software Engineer Intern

06/2021 – 08/2021

- Built a React-based geospatial intelligence dashboard and backend integrating geographic and weather APIs to accelerate tactical assessment of locations.

PROJECTS

Lighthouse Church Chatbot - <https://lighthouse-sermon-bot.matthew-peng.com/>

- Architected and deployed an end-to-end RAG platform transforming 1,500 hours across 2,028 sermon recordings into 46K+ searchable vector records through automated Whisper transcription, chunking, OpenAI embeddings, and Pinecone ingestion.
- Engineered a LangChain/LangGraph inference system combining Pinecone semantic retrieval with MySQL-backed metadata tools for structured filtering and aggregation, plus query rewriting, result validation, citation-grounded generation, and OpenAI/self-hosted Ollama model support.

EDUCATION

University of Southern California | Los Angeles, CA

M.S. Computer Science

01/2022 – 05/2023

B.S. Computer Science

08/2019 – 05/2023