

## Peifeng Wang

Tokyo, Japan | [root@wangpeifeng.com](mailto:root@wangpeifeng.com) | [github.com/peifengstudio](https://github.com/peifengstudio) | [linkedin.com/in/wangpeifeng](https://linkedin.com/in/wangpeifeng)

### SUMMARY

---

Senior Full-Stack / Lead Engineer with 10+ years building production systems in Python, with recent focus on LLM-powered products — RAG pipelines, multi-provider model routing, and AI evaluation systems. Delivered 0-to-1 products, scaled APIs to millions of daily requests, and led engineering teams of up to 10.

### HIGHLIGHTED ACHIEVEMENTS

---

- **LLM-Powered Data Platform** — Built an end-to-end financial intelligence platform with hybrid RAG retrieval (BM25 + vector + RRF), multi-provider LLM routing (Ollama / OpenAI / Claude), and LLM-as-judge evaluation. Celery + pgvector architecture with Claude prompt caching (~90% cost reduction).
- **0-to-1 Commerce Platform** — Architected and launched a full commerce platform (Rails / Next.js / Stripe), supporting **1,000+ transactions (¥10K–¥200K/order)**.
- **Warehouse Automation Patent** — Built a parametric warehouse design system generating CAD/2D/3D layouts from web inputs, eliminating manual design and resulting in an **international patent (WO2021184193A1)**.
- **Scaled Platform & Team** — Grew from **sole developer to Web Team Lead of 10 engineers at Geetest**, scaling the core B2B platform from **<10K to 200K** users.

### WORK EXPERIENCE

---

#### Full-Stack Engineer · Lunaris

Tokyo, Japan · Aug 2025 – Present

*Shopify-focused e-commerce company. Joined to build a new SaaS product — a timecard management tool currently in internal pilot.*

- Built a **real-time status dashboard** across all locations, eliminating polling and reducing system load while improving operational visibility into employee activity, implemented with **Phoenix LiveView** and **WebSockets**.
- Designed an **animated map interface** visualizing employee **check-in activity in real time** using a **2D canvas** with **Anime.js**.

#### Lead Full-Stack Engineer · Foriio

Tokyo, Japan · Aug 2024 – Jul 2025

*Japanese startup building portfolio and commerce products for creators.*

- Took ownership of **architecture, delivery priorities, and team operations** as **Lead Engineer, rebuilding the engineering team** and processes from the ground up.
- **Architected and launched a full commerce platform** (Rails / Next.js / TypeScript / Stripe), supporting **1,000+ transactions** with order values ranging from **¥10K to ¥200K** in a creator marketplace.
- **Maintained and refactored** Ruby on Rails APIs supporting **2–3 million daily requests**; resolved recurring performance bottlenecks.
- Modernized CI/CD pipelines across **10+ projects** and **40+ servers** using modular **GitHub Actions** architecture.
- **Led a team of 4 engineers** — interviews, onboarding, task allocation, and priority definition.

#### Software Engineer · Gennmu Technology

Tokyo, Japan · Jan 2023 – Jul 2024

*Tokyo-based staffing and technology firm serving major Japanese enterprises.*

- Built internal web systems with Flask and Django; delivered Salesforce Marketing Cloud solutions for enterprise clients (AU, UNITED ARROWS).

#### Full-Stack Engineer / Technical Director · Shucheng AI

Shanghai, China · Jul 2019 – Feb 2022

*Supply chain digitization company incubated by SF Express, China's largest logistics enterprise.*

- **Re-architected** the company's primary **Python-based** platform from the ground up — redesigning core systems and technical foundations.
- Architected a **warehouse design automation tool** generating **CAD, 2D, and 3D layout** models from parametric web inputs — eliminated manual design cycles for logistics PoC clients. (Patent: WO2021184193A1)
- Led full-cycle recruitment and managed a **team of 5 engineers**.

**Senior Full-Stack Engineer** · BITO Robotics Shanghai, China · Jul 2018 – May 2019

*Robotics company specializing in AGV systems; core team from Carnegie Mellon University.*

- **Rebuilt AGV central control system** (Flask / Redis / Vue.js) with a custom **WebSocket** bridge to **ROS** for **real-time bidirectional communication** with an **autonomous forklift fleet**.
- Developed a road network design tool using D3.js — constructs warehouse maps from QR code scans, allowing operators to define nodes, roads, and AGV routes via web interface.
- Led a development **team of 3 engineers** across both systems.

**Full-Stack Engineer** → **Web Team Lead** · Geetest Wuhan, China · Jun 2014 – May 2018

*Cybersecurity startup (CAPTCHA/bot-detection). Joined as the 13th employee; company scaled to ~200 employees.*

- Joined as **sole web developer** and became **Web Team Lead within one year** — ultimately led a team of 10 engineers.
- Architected the **core B2B customer platform** (user management + API key portal) scaled from under **10K to 200K** registered users; rebuilt multiple times (**Django** → **Flask + Vue.js**) to keep pace with growth.
- Delivered **10+ production web systems** across **full product lifecycle** including homepage rebuilds, **BI dashboards**, and **ML team tooling**.

## PROJECTS

---

**Personal Project** — sky-finance (LLM-powered financial intelligence platform)

- Built an end-to-end financial intelligence platform — market/news ingestion, LLM-assisted cleaning, embedding, and RAG-powered strategy analysis.
- Implemented hybrid RAG with pgvector, BM25 full-text search, HNSW vector search, and Reciprocal Rank Fusion; sentiment-bucketed retrieval ensures minority market signals reach the model.
- Built an LLM-as-a-judge evaluation pipeline comparing retrieval modes across faithfulness, coverage, and actionability; integrated multi-provider LLM routing (Ollama / OpenAI / Claude) with prompt caching (~90% cost reduction).

## TECHNICAL SKILLS

---

<b>Languages:</b>	Python, JavaScript / TypeScript, Go, Elixir, Ruby
<b>Frameworks:</b>	Django, FastAPI, React, Next.js, Phoenix (LiveView), Rails
<b>AI &amp; LLM:</b>	OpenAI API, Claude API, Ollama / Qwen, RAG, pgvector, LLM Pipelines
<b>Databases:</b>	PostgreSQL, MySQL, MongoDB, Redis
<b>DevOps:</b>	Docker, AWS, GCP, Linux, GitHub Actions, Ansible

## EDUCATION & OTHER

---

**Wuchang University of Technology** · Environmental Engineering · Sep 2007 – Jul 2011

**Languages:** Mandarin (Native) · English (Business) · Japanese (Basic)

**Patent:** WO2021184193A1 (Computer-Aided Warehouse Space Planning)