

ワンペイフン (Peifeng Wang)

東京都, 日本 | root@wangpeifeng.com | github.com/peifengstudio | linkedin.com/in/wangpeifeng

職務要約

Pythonを中心に10年以上の本番システム開発経験を持つシニアフルスタック/リードエンジニア。近年はLLMプロダクト開発 (RAGパイプライン、マルチプロバイダーモデルルーティング、AI評価システム) に注力。0→1プロダクトの立ち上げ、数百万リクエスト規模のAPI運用、最大10名のエンジニアチームのリードを経験。

主な実績

- LLM金融インテリジェンス基盤 – ハイブリッドRAG (BM25 + ベクトル + RRF)、マルチプロバイダーLLMルーティング (Ollama / OpenAI / Claude)、LLM-as-judge評価を備えた金融インテリジェンス基盤を構築。Celery + pgvector構成にClaudeプロンプトキャッシングを導入し、APIコストを約90%削減。
- 0→1コマースプラットフォーム – Rails / Next.js / Stripeを用いたフルコマースプラットフォームを設計・ローンチ。1,000件以上のトランザクション (¥10K~¥200K/件) を処理。
- 物流倉庫自動化特許 – Webインターフェースへのパラメータ入力からCAD・2D・3Dレイアウトを自動生成する倉庫設計システムを構築。手動設計を不要とし、国際特許を取得 (WO2021184193A1)。
- プラットフォームとチームのスケール – GeetestにてソロWebエンジニアからWebチームリード (10名) へ成長。コアB2Bプラットフォームのユーザー数を1万人未満から20万人へ拡大。

職務経歴

フルスタックエンジニア・Lunaris

東京都・2025年8月～現在

Shopify向けECソリューション企業。新規SaaSプロダクト (勤怠管理ツール) の開発メンバーとして参画。

- Phoenix LiveView・WebSocketsを用いたリアルタイム勤怠ステータスダッシュボードを構築。ポーリングを廃止しシステム負荷を削減。
- Anime.jsと2Dキャンバスを使用し、従業員の打刻アクティビティをリアルタイムで可視化するアニメーションマップインターフェースを設計。

リードフルスタックエンジニア・Forio

東京都・2024年8月～2025年7月

クリエイター向けポートフォリオ・コマースプロダクトを開発するスタートアップ。

- リードエンジニアとしてアーキテクチャ・デリバリー優先度・チーム運営を担い、エンジニアリング組織とプロセスをゼロから再構築。
- Rails / Next.js / TypeScript / Stripeを用いたフルコマースプラットフォームを設計・ローンチ。クリエイターマーケットプレイスにて¥10K~¥200Kの取引を1,000件以上処理。
- 1日200~300万リクエストを処理するRuby on Rails APIのメンテナンス・リファクタリング、パフォーマンスボトルネックの解消。
- モジュール型GitHub Actionsにより10以上のプロジェクト・40以上のサーバーのCI/CDパイプラインを刷新。
- 4名のエンジニアをリード (採用・オンボーディング・タスク配分・優先度定義)。

ソフトウェアエンジニア・Genmu Technology

東京都・2023年1月～2024年7月

日本の大手企業にITサービスを提供する東京のスタッフィング・テクノロジー企業。

- FlaskおよびDjangoを用いた社内向けWebシステムを構築。エンタープライズクライアント（AU、UNITED ARROWS）向けにSalesforce Marketing Cloudソリューションを提供。

フルスタックエンジニア / テクニカルディレクター・Shucheng AI 上海, 中国・2019年7月～2022年2月
中国最大の物流企業SF Expressがインキュベートしたサプライチェーンデジタル化企業。

- Pythonベースのコアプラットフォームをゼロから再アーキテクチャ。
- Webインターフェースへのパラメータ入力からCAD・2D・3Dレイアウトを自動生成する倉庫設計自動化ツールを構築。物流PoCクライアントの手動設計を廃止。（特許：WO2021184193A1）
- 採用全工程を担当し、5名のエンジニアチームをマネジメント。

シニアフルスタックエンジニア・BITO Robotics 上海, 中国・2018年7月～2019年5月
AGVシステム専門のロボティクス企業（コアチームはカーネギーメロン大学出身）。

- Flask / Redis / Vue.jsでAGV中央制御システムを再構築。自律フォークリフトとのリアルタイム双方向通信を実現するカスタムWebSocketブリッジ（ROS連携）を開発。
- D3.jsによる道路ネットワーク設計ツールを開発。QRコードスキャンから倉庫マップを構築し、ノード・道路・AGVルートをWeb上で定義可能に。
- 3名のエンジニアチームをリード。

フルスタックエンジニア → Webチームリード・Geetest 武漢, 中国・2014年6月～2018年5月
CAPTCHA・ボット検知スタートアップ。13番目の社員として入社、在籍中に約200名規模へ成長。

- 唯一のWebエンジニアとして入社し、1年でWebチームリードに昇格。最終的に10名のチームを率いた。
- コアB2Bプラットフォーム（ユーザー管理・APIキーポータル）を設計・構築。Django → Flask + Vue.jsへの複数回の再設計を経て、ユーザー数を1万未満から20万に拡張。
- ホームページ刷新・BIダッシュボード・ML向けツールなど、10以上の本番Webシステムを開発・納品。

個人プロジェクト

sky-finance — LLM活用型金融インテリジェンスプラットフォーム

- 市場・ニュースデータの取得、LLMによるデータクリーニング、エンベディング、RAG検索、金融戦略分析までを網羅したエンドツーエンドの金融インテリジェンス基盤を構築。
- pgvector・BM25全文検索・HNSWベクトル検索・Reciprocal Rank FusionによるハイブリッドRAGを実装。センチメント別バケット検索によりマイノリティシグナルをモデルコンテキストに確実に反映。
- 検索モードをFaithfulness・Coverage・Actionabilityで比較するLLM-as-judge評価パイプラインを構築。マルチプロバイダーLLMルーティング（Ollama / OpenAI / Claude）とプロンプトキャッシング（コスト～90%減）を統合。

スキル・学歴・その他

言語： Python (Django、FastAPI) 、JS/TS (React、Next.js) 、Go、Elixir (Phoenix) 、Ruby (Rails)
AI & LLM： OpenAI / Claude / Ollama、RAG、pgvector
DB： PostgreSQL、MySQL、MongoDB、Redis
DevOps： Docker、AWS、Linux、GitHub Actions
学歴： 武昌工学院・環境工学・2007年9月～2011年7月
言語： 中国語（母国語）・英語（ビジネスレベル）・日本語（基礎～日常会話）
特許： WO2021184193A1（コンピュータ支援倉庫スペース計画）