



Vũ Văn Hải

Full-stack Developer | AI-native development

TP. Hồ Chí Minh, Việt Nam haivv.se@gmail.com Điện thoại: 0869 643 629 github.com/vuhai2002 vuhai.io.vn
linkedin.com/in/haivv

TÓM TẮT

Full-stack Developer, 3+ năm kinh nghiệm, thiên backend Node.js/TypeScript + PostgreSQL. Tự dựng và vận hành toàn bộ backend cho hai nền tảng đang phục vụ kinh doanh thật (200+ và 264 endpoint): thanh toán, hóa đơn điện tử, audit log chống giả mạo, hạ tầng tự quản 12 dịch vụ. Làm việc theo hướng AI-native: điều phối 10+ AI coding agent qua orchestrator dưới cổng review của mình; tác giả một AI agent mã nguồn mở (MIT) và trợ lý AI đọc dữ liệu CRM đang phục vụ đội vận hành.

KINH NGHIỆM LÀM VIỆC

Full-stack Developer - CES Global

02/2026 - Hiện tại

TypeScript, Express 5, PostgreSQL, Drizzle ORM, React 19, TanStack, Vercel AI SDK, Docker

- Thiết kế và xây dựng toàn bộ backend nền tảng CRM + bán khóa học đang vận hành (200+ endpoint, 45 bảng PostgreSQL, 18 cron job) cùng giao diện CRM quản trị 25 màn hình; phục vụ đồng thời đội sale/kế toán, thư viện tài liệu số CES Academy và ~12 landing page.
- Tự động hóa trọn vòng đời đơn hàng từ chuyển khoản tới hóa đơn: cổng SePay (webhook đối soát tự động, chống xử lý trùng, cửa sổ hồi 24 giờ cho giao dịch về muộn) và hóa đơn điện tử MISA meInvoice phát hành ngay sau thanh toán với 3 lớp chống trùng, khớp tuyệt đối số tiền khách trả.
- Xây trợ lý AI đọc dữ liệu CRM đang phục vụ đội vận hành trên Vercel AI SDK: agent loop với 18 công cụ chỉ đọc, mọi phép tính chạy trong SQL để model không bịa số, stream câu trả lời kèm trace từng bước tra cứu; qua 4 vòng review đối kháng trước khi merge.
- Dựng panel "AI phân tích hàng ngày" trên dashboard: 10 bộ dò SQL sinh phát hiện (khóa hủy bất thường, khách mua lặp, thời điểm chào lại, nguồn không chuyển đổi...), LLM chỉ dẫn đặt và bị máy đối chiếu từng con số trong câu; mỗi thẻ dẫn thẳng tới danh sách đã lọc.
- Vận hành email marketing tới hơn 11.000 người nhận bằng engine cron cấu hình trong CSDL, gửi theo lô có cầu dao ngắt; kèm ~113 script vận hành (đối soát ngân hàng, cứu đơn, nhập liệu hàng loạt) idempotent, mặc định chạy thử.

Full-stack Developer (part-time) - LITEspace

04/2026 - Hiện tại

TypeScript, Hono 4, PostgreSQL, Drizzle ORM, BullMQ/Redis, Next.js 16, Zalo Mini App SDK

- Xây dựng toàn bộ backend nền tảng đang vận hành (264 endpoint, 50 bảng, 38 module nghiệp vụ, 6 hàng đợi nền) phục vụ đồng thời web khách hàng, CRM nội bộ và Zalo Mini App - riêng Mini App tự phát triển trọn gói.
- Thiết kế nhật ký kiểm toán chống giả mạo theo chuẩn lưu trữ chứng từ kế toán: chuỗi băm HMAC-SHA-256 tính bằng trigger trong PostgreSQL, bảng chỉ-thêm cưỡng chế ở tầng quyền, endpoint kiểm chứng toàn chuỗi, lưu trữ 3 tầng với archive khóa ghi 10 năm trên object storage.
- Giải quyết toàn bộ tranh chấp đồng thời không dùng khóa bi quan: compare-and-swap bằng UPDATE có điều kiện, chỉ mục duy nhất bộ phận làm khóa nghiệp vụ, trần nguyên tử trong WHERE cho hạn mức khuyến mãi; webhook thanh toán 6 lớp bảo vệ dùng chung một hàm commit với xác nhận thủ công.
- Mô hình hóa hệ thống giá: VAT nhiều mức tính theo nhóm thuế suất với ràng buộc CHECK ở CSDL, khuyến mãi xếp 2 tầng, chốt giá snapshot 3 giai đoạn để lịch sử đơn không bị viết lại khi đổi cấu hình; hoa hồng cộng tác viên lưu dạng snapshot.
- Dựng lớp xác thực cho ba client với ba tư thế token khác nhau (httpOnly cookie, Bearer, Bearer + persist) trên cùng một dịch vụ: xoay refresh token có phát hiện tái sử dụng và cửa sổ ân hạn 10 giây cho nhiều tab; xác định IP client từ địa chỉ TCP peer + allowlist CIDR để header giả không đầu đọc audit log và cảnh báo.

Fresher .NET Developer - 5S Enterprise Software Solutions

08/2025 - 01/2026

C#, ASP.NET (.NET Framework), SQL Server, JavaScript

- Phát triển và bảo trì các module web nghiệp vụ trên hệ thống ASP.NET (.NET Framework) + SQL Server đang phục vụ khách hàng doanh nghiệp, làm việc trực tiếp trên codebase legacy.
- Viết truy vấn SQL và logic tầng dữ liệu phức tạp giúp giảm lỗi sai lệch dữ liệu; phối hợp QA và bộ phận nghiệp vụ phân tích yêu cầu, xử lý sự cố dữ liệu và đưa bản sửa lên UAT lẫn production.

Backend Lead (đồ án tốt nghiệp) - Green Shine Trading Service

12/2024 - 05/2025

.NET 8, ASP.NET Core, EF Core, PostgreSQL, SignalR, PayOS, Firebase, Docker, GitHub Actions

- Dẫn dắt 2 kỹ sư backend trong nhóm 4 người xây nền tảng dịch vụ giặt là cho doanh nghiệp thật: kiến trúc 5 tầng, 28 controller, 5 vai trò người dùng, realtime bằng SignalR (chat, khiếu nại, theo dõi vị trí tài xế).
- Tích hợp PayOS (webhook xác minh HMAC-SHA256), Backblaze B2, Mapbox, push notification Firebase; đóng gói Docker và CI/CD GitHub Actions, giảm 35% thời gian đưa bản lên.

DỰ ÁN TIÊU BIỂU

Zalo Agent

Mã nguồn mở - MIT

TypeScript, Vercel AI SDK, Hono, React 19, SQLite FTS5, Docker

- Thiết kế và xây dựng AI agent chạy production trên Zalo: agent loop tự viết, 15 công cụ nội bộ + tool nạp động từ MCP server ngoài, nhiều tài khoản trên 2 kênh (nick cá nhân + Bot API chính thức) trong một tiến trình, dashboard 82 endpoint; ~89.000 dòng code, 359 commit.
- Xây kho tri thức RAG cục bộ: nạp txt/md/docx/xlsx/pdf, cắt đoạn theo ranh giới tự nhiên, tìm bằng SQLite FTS5 (bm25) + Reciprocal Rank Fusion, cột tìm kiếm bỏ dấu để xử đúng tiếng Việt; nguồn gán theo từng agent, mặc định đóng.
- Phòng thủ prompt injection nhiều lớp cho agent đọc tin người lạ: bọc nội dung ngoài như dữ liệu, guard SSRF chặn cả redirect và DNS rebinding, chặn rò system prompt ở đầu ra, MCP chỉ nhận HTTP Streamable với default-deny hai lớp; cố ý không xây tool tài chính.
- Kiểm soát chi phí bằng số đo thật: ngân sách ngữ cảnh theo token (đo lượt 184.835 token/8 bước), giảm token ảnh ~12.600 -> ~2.600 bằng đổi ưu tiên kích thước; 17 eval chạy model thật; trace từng bước agent lưu cục bộ để không đẩy hội thoại người dùng ra SaaS.

github.com/vuhai2002/zalo-agent

Nền tảng Streaming & Quản trị nội dung

Solo - full-stack

Next.js 16, Hono 4, PostgreSQL, Drizzle, Electron, ffmpeg, Backblaze B2, Docker

- Tự phát triển toàn bộ nền tảng streaming video/audio (web công khai + CMS trong một container Next.js qua host routing, backend Hono 81 endpoint) và app desktop Electron chạy ffmpeg native để đăng bài giảng 300 MB - 2 GB vượt trần ffmpeg-wasm của trình duyệt; quản lý gần 2 TB media trên object storage sau CDN.
- Rút thời gian tìm kiếm trong phụ đề từ 54 giây xuống ~160 ms bằng cách đổi đơn vị truy hồi sang segment cửa sổ trượt trên 1,8 triệu cue (PostgreSQL FTS + GIN), đạt Recall@1 92% / MRR 0,94 với bộ eval lặp lại được.
- Thiết kế RBAC động với 3 mức riêng tư nội dung và step-up authentication cho 9 thao tác nguy hiểm; upload thẳng lên object storage theo phiên tải ghi-trước (record-first) an toàn khi crash; SEO/GEO với JSON-LD VideoObject kèm transcript để công cụ tìm kiếm và trợ lý AI trích dẫn được.

AI Translator

Chrome Extension - 359 người dùng, 5.0 sao

JavaScript, Chrome MV3, Express 5, Docker

- Phát triển và duy trì 18 tháng tiện ích dịch AI đạt 359 người dùng, 5.0 sao/9 đánh giá trên Chrome Web Store: dịch khi hover, bôi đen, menu chuột phải, 50+ ngôn ngữ, hai đường dịch theo đánh đổi tốc độ - chất lượng.
- Xây backend proxy Express 5 giữ API key phía server, fallback tự động sang nhà cung cấp dự phòng, dịch theo lô có chia đôi khi lỗi, rate limit nhiều lớp theo IP + User-Agent + install token; deploy Docker hardened sau Cloudflare + Caddy.

chromewebstore.google.com/detail/hkelkkhcmkhhbkenoogdeomannakpilk

HẠ TẦNG & DEVOPS

- Tự vận hành 12 dịch vụ production trên VPS theo một chuẩn triển khai thống nhất: Docker multi-stage chạy non-root, cap_drop ALL, rootfs chỉ đọc, container bind loopback sau Caddy, deploy idempotent gác bằng healthcheck; chuẩn này đóng gói thành công cụ tái sử dụng.
- Hardening VPS Ubuntu 24.04 chứa dữ liệu kinh doanh thật bằng 17 script idempotent + 6 unit systemd với 10 lớp phòng thủ (firewall chỉ nhận dải IP CDN, SSH key-only đổi port, fail2ban ban leo thang, 14 tham số kernel, vá tự động + livepatch); dừng lại máy từ đầu trong 30-45 phút.
- Xây pipeline sao lưu PostgreSQL mã hóa bằng age trước khi rời máy, lưu 3 tầng trên object storage, dead-man switch báo khi backup ngừng chạy; cảnh báo thời gian thực + heartbeat hằng ngày, thiết kế để tự phát hiện khi chính lớp giám sát ngừng hoạt động.

KỸ NĂNG CHUYÊN MÔN

AI / LLM: Agentic coding, multi-agent orchestration, RAG, tool calling, MCP, Vercel AI SDK, agent evals, prompt-injection defense

Ngôn ngữ: TypeScript, JavaScript, C#, SQL, Bash

Frontend: React, Next.js, TanStack Query/Router, Tailwind CSS, Electron, Zalo Mini App

Backend: Node.js, Hono, Express, ASP.NET Core, REST APIs, SignalR, BullMQ

Cơ sở dữ liệu: PostgreSQL, SQL Server, SQLite (FTS5), Redis, Drizzle ORM, Entity Framework Core

DevOps: Docker, GitHub Actions CI/CD, Ubuntu hardening (UFW, fail2ban, systemd), Caddy, Cloudflare, S3-compatible storage

Kiểm thử: Vitest, Node.js test runner, PostgreSQL integration testing, agent evaluations

HỌC VẤN

FPT University, TP. Hồ Chí Minh - Cử nhân Kỹ thuật phần mềm (09/2020 - 05/2025). Học bằng 30% nhờ kết quả thi đầu vào.

CHỨNG CHỈ & GIẢI THƯỞNG

- Certificate of Excellence - Cuộc thi Khởi nghiệp, FPT University (2024)
- IELTS 5.5