

Wei-Han Chi

✉ whccchi@gmail.com

Production-minded Software Engineer

AI Applications · Backend Systems · Production Observability



| SUMMARY

Engineer with experience across product engineering, backend systems, and LLM applications. Strong at turning ambiguous requirements into deliverable architecture, implementation, and operations workflows. Delivered RAG and LLM services, MCP context source integrations, offline deployments, ticketing transaction systems, production migrations, observability stacks, and system handoff.

| SELECTED PROJECT

2026

Event Ticketing Platform | Production Project

Cloudflare Workers Durable Objects Prisma ORM PostgreSQL k6 Logpush DuckDB

Led the development of the platform's core order and ticketing data flow, including ticket data management, seat selection, order creation, data validation, production migration, observability, and handoff. Integrated the waiting room, lottery, and call-in flow, owned the transaction path design, concurrency load testing, and go-live support, and continued to support the team as an advisor after launch.

• Highlights:

Designed and implemented order creation, seat and ticket lock / release, and transaction validation flows to handle race conditions under concurrent purchases and prevent overselling, duplicate locks, and data mismatch.

Designed full-flow load tests with k6 covering lottery, call-in, queueing, seat selection, order creation, and data validation to verify stability and consistency during peak sales.

Investigated HTTP RTT, deployment region, and SQL transaction cost bottlenecks in a serverless architecture, then improved create-order throughput from about 20 req/s to 100+ req/s through transaction path refactoring while maintaining zero oversell.

Built ticketing data management workflows covering venue layouts, ticket types, sessions, product availability, and validation to support both back-office operations and the storefront sales flow.

Integrated a Durable Objects-based waiting room, lottery, and call-in flow, improving the handoff between queueing, admission, seat selection, and order creation under traffic spikes.

Wrote a production data migration playbook covering extraction, transformation, backfilling, validation, and rollback SQL to reduce cutover risk.

Introduced structured logging, request tracing, Logpush + DuckDB query workflows, Slack alerts, and handoff guides so the team could troubleshoot production issues independently.

| EXPERIENCE

2024/07 – 2026/01

Docubank | Solutions Engineer

NestJS(Node.js) Prisma ORM GraphQL Tanstack + React.js FastAPI (Python)

Own architecture design, LLM application services, on-premise and offline deployment SOPs, observability stacks, and product cost analysis for enterprise document workflow products, supporting the full path from requirement breakdown and system design to deployment and operations.

• Highlights:

Built type-safe APIs with NestJS, Prisma ORM, and GraphQL, and contributed to full-stack product development with React and TanStack for enterprise document workflows and AI features.

Developed a cost-aware RAG service with vector storage and similarity query APIs for semantic retrieval across multiple document sources, reducing LLM token usage by about 40%.

Unified multiple data sources including NAS, email, FTP, and user uploads into shared storage used as MCP context sources.

Delivered on-premise and offline deployments for LLM applications and backend services with Docker images, including SOPs and delivery processes for restricted network environments.

Introduced Vault for centralized secrets management, runtime configuration updates, and least-privilege access control to reduce secret exposure and operations risk.

Integrated trigger.dev for long-running AI and data workflows, moving heavy jobs off the main request path and reducing service coupling and synchronous load.

Designed and maintained multi-service GitLab CI/CD pipelines, using multi-stage builds and layer caching to reduce image size by 60% and cut CI build time from 25 minutes to 15 minutes.

Built a Grafana, Prometheus, Loki, and Promtail observability stack with Slack alerts to improve production monitoring and incident investigation.

Planned service topology, deployment patterns, and network connectivity for on-premise and offline product delivery.

Analyzed AWS, GCP, and third-party service costs, then combined k6 load-test results with usage projections to estimate log economics for pricing and cost optimization.

2022/09 – 2024/06

CYBERBIZ | Lead Engineer

FastAPI (Python) Next.js (React.js) Ruby on Rails

Led APP Market product planning and execution, PR reviews, and custom feature development.

• Highlights:

Led backend and frontend technology selection and system architecture for APP Market, owned backend APIs and Admin Panel implementation, and coordinated with DevOps and frontend teammates through delivery.

Designed and implemented a plugin-like app specification and installation flow, extending validation without breaking existing APIs so both new and legacy customers could be supported together.

Built internal wiki documentation to help new teammates ramp up and improve cross-team communication.

Served as Scrum Master for a 3-person team, working with PMs to refine requirements and split work, while also providing engineering guidance for communication with external vendors.

Delivered multiple custom features including back-office and storefront SSO, custom customer fields, enhanced Google conversion tracking through GTM, and member data masking for enterprise clients.

2020/12 – 2022/06

OwlTing Group | Software Engineer

Laravel (PHP) FastAPI (Python) Vue.js Drone (CI/CD) Docker

Worked in the blockchain lab, building products including an IoT Gateway, OwlTing NFT Marketplace, and a supply-chain traceability system across gateway services, frontend interfaces, API documentation, third-party integrations, and production operations.

• Highlights:

Rebuilt the IoT Gateway system with device grouping, FSM-based state management, blockchain fault tolerance, and a data notification management interface, improving device management, data traceability, and on-chain workflow stability. The new and legacy systems ran in parallel for a period to validate data and support a gradual handoff, with zero data loss during the transition.

Built the notification management interface and API documentation to support communication with PMs, external vendors, and third-party integrations, reducing rollout and maintenance overhead for the IoT Gateway.

Owned the order, payment, invoice, and on-chain transaction flows for OwlTing NFT Marketplace, implementing transaction replay, rollback, and compensation logic.

Maintained the Taiwan forest-products traceability system, handling full-stack development, requirement interviews, third-party integrations, and production incident response.

2018/11 – 2020/12

Commonwealth Magazine Group | Senior Engineer

Laravel (PHP) Flask/Django (Python) Koa (Node.js) GCP

Worked as a backend engineer in the IT department, building product features, internal order and subscription flows, media APIs, content tooling, and productionized AI systems.

• Highlights:

Planned the system architecture, deployment, frontend-backend integration, and ETL for the 2020 Taiwan presidential election special project, which won first place in the SOPA 2020 Excellence in Data Visualization award.

Built internal order and subscription flows with subscription state management, renewal logic, and branching scenarios to support subscription products and operations.

Created a user tagging system that turned third-party log data into a user-id-based tagging index and used MongoDB read/write separation to support segmentation, digital marketing, and content product needs.

Developed an article keyword system with a Django-based retraining flow, admin workflows, API integrations, and model update processes to support content tagging, search, and recommendation use cases.

Collaborated on productionizing an AI article recommendation system by working with AI consultants to expose model capabilities via APIs and designing a consistent recommendation interface consumed by multiple products, while also supporting data collection for later model tuning.

Built a dynamic thumbnail and campaign service that included Google Sheets-based whitelist management, game caching, state management, and an achievement system to support editorial operations and fast delivery for project-based products.

2014/04 – 2018/11 | Earlier Experience

• 掌櫃智慧生活 | PHP Engineer:

Node.js MQTT Laravel

Built a new admin system and designed MQTT-based counter status monitoring to support real-time device state tracking.

• 碩網資訊 | Software Engineer:

Express Python PHP

Worked on government open data platforms, a Taipei City Government LINE Bot, finance-industry POCs, and product environment consulting. Handled product development, deployment, and API integrations for public-sector and finance clients.

• Aurora Group | Engineer:

LAMP PHP

Developed and maintained Aurora office cloud e-commerce services, and planned local data-center migration SOPs and validation workflows to ensure customer data integrity.

| EDUCATION

2008-09 – 2012-09

National Kaohsiung University of Applied Sciences

B.B.A. in Information Management