Jay Zeng

I lead international engineering teams building platforms end-to-end across regulated industries — finance and healthcare

Shanghai, China · +86-17721216772 | jay.zeng@hotmail.com
linkedin.com/in/hwajay | jay.obolai.hk | obolai.hk

From clinical imaging at AGFA Healthcare and retail banking at Standard Chartered China to VP at crypto.com — hands-on throughout: coding, code review, system design and architecture review; strong opinions on clean code, TDD, spec-driven development; long-time Agile practitioner, from Scrum to pods. Led 30 engineers and QAs across six locations building: a real-time US-equity market-data platform; a token-price platform and a digital-asset staking platform; a tokenized-securities (RWA) business through regulatory review to production; and applied-AI platforms on the group's internal AI gateway. Drove AI-assisted development to org-wide adoption with engineering guardrails.

Technical Stack

Backend & Distributed Systems: Java (Spring Boot, Spring WebFlux), Go, microservices & event-driven architecture, messaging (Kafka, RabbitMQ), PostgreSQL, MongoDB, ClickHouse, Redis, REST / WebSocket / GraphQL APIs

Cloud & Infra: AWS (EKS/Kubernetes, Bedrock), GCP (Docker Compose, KMS), Helm, CI/CD, DevSecOps, observability (Datadog, Sumo Logic, Grafana, OpenSearch, Prometheus)

AI / Agents (hands-on): LLM pipelines & multi-model agents, multi-agent orchestration (LangGraph), RAG (embeddings + Postgres pgvector), LLM evaluation & routing, human-in-the-loop control design, AI governance (fail-closed controls), full-stack LLM application engineering (FastAPI services, Next.js operator tooling), AI coding tools (Claude Code, Cursor, Augment)

AI Models: DeepSeek 4+, GLM 5+, Anthropic Sonnet / Opus 4.8

Languages: Java, Go, TypeScript, Python

Engineering Leadership: people management (multi-team multi-site org), developing engineers to high engineering bars, Agile practices, performance & calibration, budget management, market-data vendor relationships, governance, risk & regulatory delivery

Domains: retail & institutional banking platforms, capital-markets data infrastructure, real-time market data, digital-asset platforms (RWA tokenization, staking), clinical imaging platforms (DICOM/HL7/FHIR)

crypto.com

Nov 2020 – Jul 2026 · 5.5 yrs

VP / Engineering Lead — Shanghai

Global financial platform serving 100M+ users; fully licensed and regulated across multiple jurisdictions

- Led 30 engineers and QAs in AI-platform, RWA & digital-asset staking, market-data, NFT, and finance-transformation automation teams across six locations (Taiwan, Shenzhen, Shanghai, Singapore, Hong Kong, London) — owning system design, architecture reviews, and code review across the portfolio, alongside headcount planning, performance management & calibration, pay/reward, and budgets
- Established the tokenized-securities (RWA) business 0-to-1 — on-chain tokenization in Go for 1,500+ US equities serving the App's 100M+ registered users, owning the full lifecycle from architecture through regulatory review to production

- Drove AI-assisted software development to 100% adoption across the engineering organization — 12+ months of tooling rollout (Claude Code, Cursor, Augment) within the group's org-wide AI-SDLC (AIDLC) transformation — setting the engineering guardrails for model evaluation, cost governance, and code ownership
- Led the team that built AI-powered applications on the group's internal AI gateway — including Augur, an LLM + RAG market-intelligence platform co-created with the crypto.com Capital team, serving investment research, CRM client engagement, and the US-equity RWA business
- Managed market-data vendor relationships (ICE, CoinGecko, CoinMarketCap)

Digital-Asset Staking Platform — Java Microservices
Designed and operated the platform's key services in Spring Boot MVC — the Staking API Gateway and the Staking Core Service — serving the crypto.com App, Exchange, internal trading, and Visa Card: daily staking booking and rebalancing, staking and payback processing both on-book and on-chain.

- **Staking API Gateway** — a lightweight intake layer: request validation and persistence, nothing else (native JAAS authentication, Redis auth cache, PostgreSQL)
- **Staking Core Service** — a self-designed Actor model: each shard — one instrument × one user-entity group (US / Canada / ROW / Card / IRA) — runs a single-threaded task loop fed by schedulers; for any instrument exactly one task is ever in flight — no locking, no concurrent-write races, data flow safe by construction

The daily rebalance nets the day's bookings and settles on-chain — staking surplus, unstaking shortfall. $100M AUM across 20+ assets.

Tech Stack: Java (Spring Boot MVC) · PostgreSQL · Redis · AWS EKS · Helm

US Stock Market Data Platform — Go · 0-to-1
Architected in Go, end to end:

- **Ingestion** — private lines into ICE's CDP; a dedicated big-memory edge node runs the low-level socket protocol over licensed Tape A/B/C and NASDAQ feeds — peak ingress 1.4M+ msgs/sec
- **Streaming backbone** — raw ticks into Kafka topics partitioned by instrument and source; a ClickHouse consumer archives every raw message for long-term query
- **Real-time processing** — source-grouped in-memory workers on dual schedulers: a 5-second scheduler flushing worker cache to Redis and publishing enriched updates to per-source topics, a 1-minute scheduler re-publishing Redis state so clients never see a stale price
- **Metadata & history** — PostgreSQL holds instrument metadata, refreshed hourly via commands back to the CDP; a companion Go service computes multi-interval candlesticks from ICE's historical endpoints
- **Asset universe** — asset lists combined from ICE and Apex (the US equity broker), managing each stock's price reporting and tradable status

Powers equity, perpetual-contract, and tokenized-stock trading.

Tech Stack: Go · Kafka · ClickHouse · Redis · PostgreSQL

Digital Asset Market-Data Platform — Java / Spring Reactive · 0-to-1
Led three core services, each owning one layer of the platform — price computation, distribution, and canonical token identity:

- **Price engine** — aggregated multi-source CEX and DEX tickers into real-time volume-weighted average prices (VWAP) for 2,000+ tokens: pairs standardized via base-currency routing, stale quotes dropped by liveness checks, spikes filtered with MAD and dynamic fallback thresholds, gated by listing breadth (7+ exchanges for the full outlier-filtered method, a dedicated thin-source path below)

- **Price provider & price writer** — the distribution path: the price provider consolidated sources by defined rules and published price reports to RabbitMQ routing keys; the price writer joined them with token metadata, persisted to a MongoDB replica set, and exposed REST and WebSocket read APIs to client services
- **Metadata service** — owned the canonical token identity across a universe of 20,000+ tokens: a single unique ID each, mapped to CoinGecko and CoinMarketCap IDs so vendor price feeds matched the right instrument. Maintained symbol, name, on-chain decimals, logo, and EEA compliance attributes — including the MiCA-required legal documents surfaced on the market website's token pages. Its ticker mappings drove the price engine's source selection and every serving-layer service attached its metadata

The platform is the source of truth for token data and market prices across the crypto.com ecosystem — powering token listing, the RWA business, financial reporting, and trade-quotation gating.

Tech Stack: Java (Spring WebFlux, Spring MVC) · RabbitMQ · MongoDB

Citadelle — Autonomous AI Agent System
Started at crypto.com with the head of Yield — prototyped through to a staging deployment — then rewritten from scratch after leaving and continued as an independent product, now testing with real funds. An end-to-end autonomous on-chain capital-allocation system: a 15-step multi-agent pipeline (LangGraph) running the complete investment loop from opportunity discovery and Bull/Bear/Judge multi-LLM adversarial debate to quantitative optimization and on-chain execution. Every LLM decision is wrapped in deterministic, fail-closed controls:

- an 8-layer concentration firewall
- scenario-and-rubric eval harnesses with nightly LLM evaluation
- degradation that never defaults to approval
- human-in-the-loop gates replaying immutable route bundles
- KMS signing isolated behind a dedicated executor service

A working blueprint for adopting autonomous AI agents.

Tech Stack: Python 3.11 / FastAPI · TypeScript / Next.js ops dashboard · PostgreSQL + pgvector · DeepSeek with GLM 5.3 fallback · Docker Compose on GCP · GCP KMS · Prometheus

Augur — LLM Market Intelligence
Cross-team AI innovation co-created with the crypto.com Capital team — the opportunity identified in-house, not from a given spec. A 5-minute news pipeline on the group's AI gateway (model usage control and cost monitoring, AWS Bedrock) classifies items by event type and materiality, with a deterministic scoring framework routing critical news into a real-time signal queue; daily RAG synthesis produces retail-ready narratives and sector outlooks for Capital's investment research, CRM's client push, and the newly launched US-equity RWA offering — with on-platform prompt tuning and side-by-side model-output comparison.

Tech Stack: TypeScript · Python · AWS Bedrock (group AI gateway) · LLM + RAG · PostgreSQL + pgvector

AI-Assisted Engineering Adoption
Rolled out AI-assisted development to 100% of the engineering organization — leading as a daily hands-on practitioner. Established guardrails for model evaluation, cost governance, and code-ownership review — countering over-reliance risks where reduced code familiarity introduced defects. Cycle-time gains validated through compliance review.

Tech Stack: Claude Code · Cursor · Augment

Standard Chartered

Jan 2019 – Nov 2020 · 2 yrs

Delivery Lead, Retail Banking — China (Shanghai)

- Led delivery of RB CDD Review — Standard Chartered China's first containerized application, digitizing retail-banking compliance (KYC/CDD review) workflows — in partnership with the China country technology management team and the CDD team
- Owned SCB China's retail-banking delivery portfolio — Java / Spring Boot microservices delivery with teams across Shanghai and Tianjin (BAU, mandatory & regulatory change, new initiatives, platform integration)
- Initiated the Shanghai API Banking engineering team build-out — stood up hiring and knowledge transfer from the Singapore core team
- Led partnership and vendor evaluation with Alipay, Ctrip Finance, and JD Finance — scoped product-listing feasibility and integration architecture for SCB retail products on third-party platforms (paused due to regulatory change)

AGFA Healthcare

Jun 2010 – Dec 2018 · 8.5 yrs

Team Lead & Technical Development Manager — Shanghai

- Led one of four scrum teams at AGFA's Shanghai R&D site, and served as site technical manager — the technical bridge to global R&D hubs in Waterloo, Ghent, and Vienna, jointly driving the evolution of the Enterprise Imaging Platform
- Led the team that delivered the platform's rule engine (Drools) and enterprise image-transfer capabilities — powering rule-directed clinical workflows, image routing and prefetching, report distribution, and storage-lifecycle management
- Full-stack engineering leadership for the Enterprise Imaging Platform — large-scale Java-based medical imaging with distributed hub-proxy architecture and high-availability deployment; domain depth across DICOM, HL7, IHE, FHIR
- Drove DevOps adoption across the Shanghai site — established CI/CD pipelines, regression testing workflows, and release management practices

Earlier Career

IWT Solutions — Consultant, Hong Kong (2009–2010): Led design and development of Internet Booking Engine for flight ticket search and fare management. Client-facing role with customers in Malaysia.

Perficient, Inc. — Sr. IT Consultant, United States (2005–2008): Led development teams on BPM-based medical equipment order processing (Apria Healthcare) and inventory control application (Rent-A-Center).

Education

MSc in Information Technology — Hong Kong University of Science and Technology (HKUST), Hong Kong SAR, 2008–2009

BEng in Software Engineering — Zhejiang University of Technology, Hangzhou, 2001–2005

Selected Writing

Published writing at jay.obolai.hk/writing on AI transformation, AI leadership, AI model economics, and regulated digital-asset infrastructure.