

Robin Rahman

Kuala Lumpur, Malaysia • rob@robinrahman.pro • [Website](#) • [GitHub](#) • [LinkedIn](#)

Professional Summary

Full-Stack AI Engineer with 8 years across e-commerce, blockchain, and production AI systems. Specializes in conversational AI, RAG pipelines, and agentic workflows, with hands-on depth in Medusa.js v1 and v2. Backend-leaning (Python, FastAPI, Node.js) with strong Next.js and Tailwind frontend capability and a Figma-to-deployment workflow. Uses Claude Code, Cursor, and Codex daily.

Technical Skills

- **AI & Generative:** LangChain, LangGraph, LlamaIndex, pgvector, Pinecone, Qdrant, OpenAI API, Anthropic Claude API, Vercel AI SDK, Stable Diffusion, HuggingFace, fine-tuning (GPT-3.5, Mistral, LLaMA)
- **Backend:** Python, FastAPI, SQLAlchemy, Node.js, NestJS, Payload CMS, BullMQ, Golang, tRPC, GraphQL, Redis, PostgreSQL, MongoDB, Prisma, Drizzle ORM, Docker
- **Frontend:** Next.js, React, TypeScript, Tailwind CSS, shadcn/ui, Jotai, Zustand, Figma, React Native, Expo
- **E-Commerce:** Medusa.js, Mercur, Shopify / Liquid, WooCommerce, Stripe
- **Blockchain / Web3:** Solidity, Hardhat, ERC-721 / ERC-20, wagmi, viem, The Graph, IPFS, Foundry
- **Cloud & Infra:** AWS (EC2, RDS, S3, CloudFront, Lambda), Railway, Vercel, Supabase, GitHub Actions
- **Testing & QA:** Jest, Vitest, Playwright, Cypress, Postman, k6

Work Experience

R² Consulting (Nov 2020 - Present)

- Full Stack AI Engineer

Mar 2023 – Present

- Deployed a production RAG chatbot (pgvector, reranking, SSE streaming) for a B2B SaaS client, cutting tier-1 support tickets by ~40% and response time from 4 hours to under 8 seconds; paired with a 200-question eval harness for automated weekly regression testing.
- Engineered a LangGraph agentic system for a D2C brand handling order lookups, returns, and shipping via internal API tool calls, sustaining 500+ concurrent AWS sessions without latency degradation.
- Built a LlamaIndex document pipeline for a legal client, cutting irrelevant retrieved context by ~50% over LangChain through native PDF parsing and node-level retrieval.
- Integrated Stable Diffusion via HuggingFace pipelines to generate product visuals and marketing assets for a D2C client, reducing creative turnaround from days to minutes.
- Delivered streaming AI chat UIs with Vercel AI SDK and Next.js across multiple clients, supporting real-time token streaming and seamless OpenAI/Claude switching without frontend rebuilds.
- Implemented Stripe usage-based billing for multiple AI SaaS products covering metered token consumption, subscription tiers, and Connect payouts, eliminating manual invoicing at scale.

- Full Stack Web3 Developer

Jan 2021 – Present

- Launched a full NFT minting platform for a 10k generative collection with an ERC-721 Merkle tree allowlist and gas-optimised batch minting (~28% lower per-token gas), sustaining 2,400 mints in the first 90 minutes without downtime.
- Developed a multi-wallet DeFi dashboard aggregating Uniswap, Aave, and Compound positions via The Graph and RPC calls, with Redis TTL caching cutting infrastructure costs by ~60%.
- Integrated GPT-3.5 to generate plain-English daily portfolio summaries from normalised on-chain data, making DeFi positions interpretable for non-technical investors.
- Architected an ERC-20 token launchpad with cliff/linear vesting, KYC allowlist gating, and a Slither-audited mainnet deploy, with a Next.js admin dashboard for non-technical pool management.

- E-Commerce Engineer

Nov 2020 – Present

- Migrated a production Medusa v1 store to v2 by rebuilding all custom plugins to the new module architecture behind feature flags and integrating Payload CMS as a headless content layer, achieving zero-downtime cutover at the October 2024 release.
- Eliminated zero-result search failures on a Medusa v2 storefront by replacing keyword search with semantic matching, directly improving product discoverability for intent-driven queries.
- Grew AOV by 18% on a Mercur multi-vendor marketplace by connecting shoppers to live inventory, vendor ratings, and delivery estimates via a real-time assistant, validated over a 6-week A/B test.

- Spearheaded an early Medusa.js v1 production migration for a DTC apparel client off WooCommerce, delivering custom multi-currency pricing and loyalty plugins alongside a zero-downtime ETL of 4,200 products and full order history.
- Automated inventory reconciliation for a regional grocery chain via Python CSV-to-WooCommerce pipelines, saving ~6 hours of manual work weekly across a multi-location platform with per-store inventory and region-specific pricing.

InvoZone · Software Engineer

Feb 2019 – Jun 2020

- Built full-stack web applications across multiple client projects using the MERN stack, delivering REST APIs and responsive React interfaces from Figma designs to production.
- Developed cross-platform mobile features using React Native for a client-facing consumer app, covering UI components, API integration, and basic push notification handling.
- Contributed to Golang microservices for a high-throughput backend system, gaining early exposure to statically-typed service architecture alongside Node.js API development.

Digital Commerce Agency · Full-Stack Developer

Oct 2017 – Feb 2019

- Rebuilt a fashion retailer's Shopify storefront with dynamic filtering, size guides, and a cart drawer, cutting bounce rate from 58% to 41% and eliminating oversell issues via real-time Node.js ERP inventory sync.
- Delivered custom Shopify Liquid themes and WooCommerce PHP plugins across multiple retail clients, integrating Stripe, PayPal, and local Malaysian payment gateways.
- Introduced Jest and established the team's first structured unit testing standard, adopted as a baseline across all subsequent projects.

Education

B.S. Computer Science, Universiti Teknologi Malaysia

2017