

Danish Shafi

London, UK · danish.shafi@live.co.uk · 07581 300 429
Portfolio: builtbydan.com · GitHub: github.com/builtbydansk · LinkedIn: linkedin.com/in/danishshafi

PROFILE

Early-career Full-Stack / Product Engineer with ~1.5 years' commercial experience building B2B HR SaaS products with TypeScript, Node.js and React/Next.js. Now building backend systems, LLM evaluation infrastructure and agentic AI, with a focus on end-to-end product ownership. Targeting Software, Product, Backend and AI Product Engineering roles.

TECHNICAL SKILLS

Languages: TypeScript, JavaScript, SQL, Python, Java
Backend: Node.js, Hono, Fastify, Express, REST/GraphQL, WebSockets, PostgreSQL, Prisma
Frontend: React, Next.js, Tailwind CSS
Systems: Event-driven architecture, Docker, k6
AI Tooling: Cursor, Codex, Claude, Grok, Antigravity, OpenAI API, LangSmith, Tavily, Runware
AI / LLM Engineering: LLM evaluations, evaluation harnesses, agentic workflows, multi-agent orchestration, tool calling, structured outputs

COMMERCIAL EXPERIENCE

Independent Software Engineer May 2025 – Present

- Building independent product and systems projects spanning backend architecture, data modelling, APIs, UI and LLM evaluation tooling.
- Experimented with multi-agent orchestration using specialised GPT Builder and Claude Evaluator agents sharing a common sandbox, with a root agent coordinating iterative build/evaluate cycles.

Software Developer | Equitas.ai — B2B HR AI SaaS Oct 2023 – Apr 2025

- Shipped core customer-facing features in React, TypeScript and REST/GraphQL APIs in an agile product team.
- Refactored legacy UI into reusable design-system primitives, reducing tech debt and frontend bundle size.
- Built Chart.js data dashboards and resolved production defects across UI and API layers.

FEATURED ENGINEERING PROJECTS

Commander — Incident Management Platform
TypeScript, Hono, Node.js, React/Next.js, PostgreSQL

- Designed an incident-management backend modelling SEV1–SEV4 incidents, role-based workflows, assignment/reassignment and explicit ownership across Dispatcher, Officer and Supervisor roles.
- Modelled an append-only incident timeline as the source of truth, preserving auditable state transitions, ownership transfers, on-call overrides and structured handovers.
- Developing the system around PostgreSQL transactions and concurrency-safe mutations, while exploring idempotency, queues/workers, caching and observability.

Trace — LLM Evaluation Harness
TypeScript, PostgreSQL, React, LLM Evals, Agent Trajectories

- Building an evaluation harness for ingesting, normalising and analysing LLM/agent evaluation runs, exposing trajectories, grader outputs, scores and failure modes across experiments.
- Designing a TypeScript/PostgreSQL evaluation pipeline with runtime validation, structured error handling and asynchronous ingestion for reliable processing of evaluation runs.
- Developing analysis tooling for comparing model/agent behaviour across datasets and runs, informed by work on sampling, dataset auditing, reproducibility and rubric leakage.

Replay — Verifiable AI Decision-Replay System
TypeScript, Next.js, Qwen3-8B, Supabase, Vercel/Railway

- Architected a multi-agent A&E decision-replay system with an Investigation Agent and verifiable decision/event receipts to reconstruct and replay AI decision processes.
- Shipped the system on Supabase with Qwen3-8B agents and deployed across Vercel/Railway for the tokens& Multi-Agents Hackathon.

Workshop — Automotive Operations & Workflow SaaS
TypeScript, Next.js, Express, Supabase (PostgreSQL), Tailwind CSS

- Architected a domain-driven vehicle lifecycle platform with Repository/Service boundaries, relational schemas and audit logging.
- Implemented workflow state machines to enforce valid operational transitions across independent workshop processes.

HACKATHONS & AWARDS

- Aptura RSI Hackathon — Top 5, 2026 — Worked with an LLM evaluation harness across health, banking/customer-service, quantitative-reasoning and STEM tasks, analysing agent trajectories, grader outputs and failure modes.
- tokens& Multi-Agents Hackathon — Winner Best Use of Cursor, 2026 — Built Replay, a verifiable multi-agent decision-replay system.

EDUCATION

MSc in Software Development | Queen's University Belfast 2024 – 2027
PGCert in Computer Science | University of York 2023