Stephen Monslow

Senior Software Engineer · Python · FastAPI · React · Elixir/BEAM · GCP · AI / LLM / RAG · Application Security

London, UK · m-onz@use.startmail.com · linkedin.com/in/stephen-monslow-94a1a4390 · github.com/m-onz · m-onz.net

PROFILE

Full-stack software engineer with 14 years' experience delivering production systems across regulated finance, fintech, aviation and embedded hardware. Specialises in Python and FastAPI backends, React and React Native frontends, and Elixir/BEAM distributed systems, deployed on Google Cloud and AWS. Builds applied-AI products end to end — LLM applications, agentic RAG and vector search — with broad machine-learning experience spanning computer vision, reinforcement learning, NLP and neuroevolution. Offensive-security researcher and bug-bounty hunter, bringing a secure-by-design approach to everything from architecture through to cloud operations and deployment.

AI & MACHINE LEARNING

Fourteen years of production systems work now pointed at AI — I build LLM applications and ML systems end to end, from classical computer vision and evolutionary algorithms to production agentic RAG, and I ship them into real products.

- **Agentic RAG in production (QDEX)** — a lending-intelligence assistant built on an LLM function-calling agent loop over a five-lane hybrid retriever (full-text, vector, prefix, graph and sibling lanes) fused with Reciprocal Rank Fusion and adaptive query classification; pgvector embedding store, sentence-transformers (MiniLM, 384-dim) embeddings, token-budget context assembly. Python/FastAPI.
- **Retrieval & vector search** — vector databases (pgvector, Pinecone, Weaviate, Qdrant), embedding models, hybrid dense/sparse retrieval, chunking and corpus-ingestion pipelines (*substrate*: 141 mixed-format documents → a clean, tagged markdown corpus with LLM-driven classification).
- **LLM apps in Python and Elixir** — agent orchestration with LangChain, CrewAI and LlamaIndex; providers across Anthropic (Claude), OpenAI, Google Gemini and local open-weight models (Ollama / llama.cpp); function/tool calling, structured output, prompt and context engineering.
- **Machine learning** — computer vision (OpenCV, Keras), supervised and unsupervised learning (scikit-learn), reinforcement learning and Q-learning, neuroevolution (authored **Bardo**, a topology-and-weight-evolving neural-network library published on Hex), spiking neural networks, online/streaming learning (River ML).
- **Applied ML research (Hedgehog AI)** — novel independent R&D in online learning, reinforcement learning, spiking neural networks and neuroevolution-driven algorithmic-trading systems.
- **Deep learning & scientific computing** — PyTorch, TensorFlow/Keras, Hugging Face Transformers; NumPy, pandas, SciPy.
- **ML Ops & serving** — containerised model serving and inference APIs, Docker, Kubernetes, GCP; CI/CD for ML workloads; evaluation and data-quality thinking from RLHF / human-data pipelines.

CORE SKILLS

- **Languages:** Python, Elixir, Erlang, Rust, JavaScript/TypeScript, Go, C, PHP
- **Backend & frontend:** FastAPI, React, React Native + Expo (iOS/Android/Web), EAS, Phoenix LiveView, Ash Framework, NumPy
- **AI & ML:** LLM apps & agentic RAG, LangChain, CrewAI, LlamaIndex, Anthropic/OpenAI/Gemini, Ollama, vector DBs (pgvector, Pinecone, Weaviate, Qdrant), sentence-transformers, PyTorch, TensorFlow/Keras, Hugging Face Transformers, scikit-learn, River ML, pandas/SciPy, OpenCV, neuroevolution, reinforcement/Q-learning, MLOps & model serving
- **Cloud & DevOps:** Google Cloud Platform, AWS, Docker, Kubernetes, Terraform, CI/CD (Jenkins, GitHub Actions), infrastructure-as-code and deployment automation
- **Systems & Linux:** Debian, Fedora, QubesOS, FreeBSD/OpenBSD, Linux/BSD system administration, system hardening
- **Security:** Web application security (offensive and defensive), vulnerability research, threat modelling, secure-by-design, OWASP, authorisation-scoping remediation (IDOR/BOLA), DevSecOps
- **Architecture:** Distributed systems, fault tolerance, high availability, microservices
- **Embedded & cyber-physical:** Espruino firmware, Arduino, computer vision, machine listening
- **Analytics & tagging:** Adobe Analytics, Tealium, tag management, analytics instrumentation and validation (web, iOS, Android)
- **Domain:** Mortgages and lending, retail banking, FX and payments, credit data

EXPERIENCE

Senior Software Engineer — QDEX Dec 2025 – Present

Greater London · Hybrid

- Sole engineer responsible for the full stack end to end: backend (Python/FastAPI), frontend (React), cloud operations (GCP) and application security.
- Own and evolve a pre-submission lending-intelligence platform (100k+ lines) — inherited the codebase and lead its development, including new feature delivery such as a purpose-built CRM layered on top of the system.
- Designed, built and shipped the platform's AI assistant — a production **agentic RAG** system answering brokers' lending-criteria questions over a curated knowledge base, including the document-ingestion pipeline that feeds it (architecture under *AI & Machine Learning*).
- Run security review of our own attack surface. Found and remediated publicly leaked test credentials, source-code exposure via a public test-coverage artefact, and authorisation-scoping gaps across the platform (IDOR/BOLA classes) — audited and fixed at the internal-scoping layer.
- Built and maintain the platform behind QDEX's move to make mortgage sourcing permanently free for every UK broker.

Senior Software Engineer (Contract) — Hedgehog AI Ltd Jun 2025 – Nov 2025

Remote

- Independent R&D across AI/ML, data and mobile — delivering working research prototypes in Python, NumPy and React Native.
- Novel machine-learning research: online/streaming learning, reinforcement learning, spiking neural networks and neuroevolution.
- Built complex algorithmic-trading experiments — online time-series prediction and neuroevolution-driven strategy search, with latency-aware execution.
- Designed and shipped **Cutlist**, a cross-platform React Native + Expo app (iOS, Android, Web) backed by an Elixir/Phoenix API — real-time sync over Phoenix Channels, RevenueCat subscriptions, full store submission and GDPR data handling.

Software Engineer — Lloyds Banking Group Apr 2020 – May 2025

London · Hybrid

- Built the tooling that made analytics instrumentation verifiable across all Lloyds brands (web, iOS, Android; Tealium + Adobe Analytics), in an environment where team self-tagging had produced drift and fudges.
- **HAR-file validator:** developers record a browser session, upload the HAR, and the tool decodes the analytics beacons and validates them against specification. Verification made asynchronous and portable — no login, no live environment — because the test estates were flaky and sometimes unreachable.
- **Automated deviation detection:** experiments that scanned real sessions against the original tagging specifications and flagged drift automatically, before it poisoned the data.
- Shipped operational and testing tools as **Confluence page-apps** — the only deployment surface available inside a locked-down corporate network. Adopted by non-technical team members and in daily use when I left.
- Full-stack delivery and architecture within enterprise-scale engineering, governance and release processes.

Backend Software Engineer — WeSwap May 2019 – Apr 2020

London

- Backend engineering for a consumer FX and travel-money platform: distributed services and Jenkins CI/CD in a regulated money-movement environment.

Contract Software Engineer — tx11 Jul 2017 – Apr 2019

London

- Contract engineering across distributed systems, backend services and automated build/deployment pipelines (AWS, Node.js, Angular, web components).

Earlier Roles

- **Principal Backend Software Engineer** — iContract, London · Feb 2017 – Jun 2017 · Backend and cloud services.
- **Senior Developer** — Experian, London · Jun 2016 – Jan 2017 · Software architecture and CI/CD at a global credit reference agency.
- **Senior Developer** — Wattbike, Nottingham · Apr 2015 – May 2016 · Software architecture and security for connected fitness products.

Co-founder & Engineer — Kielowatts Ltd (turning-target systems) Mar 2014 – Apr 2015

Lincoln

- Co-founded (with my brother) and built servo-controlled outdoor turning-target systems end to end: embedded firmware (Espruino on microcontrollers, GSM modem, SD storage, dedicated WiFi access point), a field-node mesh and a web control plane.
- Later extended the system with camera-based bullet-hole detection and distance-from-centre measurement via calibrated regression and SVM classification.

Lead Front-End Developer — Bytron Aerospace Aug 2012 – Sep 2014

Humberside

- Built aviation flight-planning software deployed inside UK airports; on-call for live airport operational systems (remote Linux operations).
- Led the push to test-driven development on Symfony and, as lead front-end architect, built custom real-time weather-visualisation widgets (Kendo UI) for operational flight-planning workflows.
- Trained by senior engineers on complex aviation systems — where availability is the product.

OPEN SOURCE

- **Martha** — contributor to an open-source agentic *software factory*: a crew of coding agents that plan, build, review and release software through event-sourced, auditable hand-offs (Node/TypeScript, Claude Code host adapter, published on npm). Part of the Macula ecosystem (macula.io · github.com/macula-io/martha).
- **Bardo** — authored and published a neuroevolution (TWEANN) library for Elixir on Hex (github.com/hibernatus-hacker/bardo · hex.pm/packages/bardo).

CERTIFICATIONS & TRAINING

- **Web & API security:** PortSwigger Web Security Academy, APIsec University, eLearnSecurity
- Active bug-bounty hunter and independent offensive-security researcher

EDUCATION

- **HND, Music Technology** — Anglia Ruskin University (real-time audio systems — the origin of my C/DSP and embedded work)
- **4 A Levels** — The Priory LSST