Ángel Gabriel Zeledón Fernández

DATA SCIENTIST & MLOPS ENGINEER — SAN JOSÉ, COSTA RICA

angel.zeledon.fernandez@gmail.com · +506 7186-2820 · linkedin.com/in/angel-zeledon-data-science · github.com/Angel-Zeledon

PROFESSIONAL SUMMARY

Data Scientist & MLOps Engineer with 4+ years building end-to-end forecasting systems, supply-chain optimisation and enterprise agentic AI — turning operational data into decisions for clients including Walmart, HP, Philips and Cisco. Cloud-native across GCP, Azure and AWS. Since June 2026 my main work has been my own two platforms — StockAI, an inventory and purchasing SaaS, and WorkforceAI, where teams run AI agents under spend caps and human approval — alongside part-time ML consulting on betting-risk models. Driven by mathematics, science, fintech and quantitative finance, and by the technology that comes next. I also build and publish research — a neural-interface layer, a Rust genomics counter, a small compiled language, a paper on liquid democracy and a 2,080-page mathematics course — each published with its limits stated next to its results.

CORE TECHNICAL SKILLS

MACHINE LEARNING & FORECASTING

Demand & time-series forecasting, predictive modelling, supply-chain analytics, inventory optimisation, feature engineering & selection, hyperparameter tuning, cross-validation, walk-forward evaluation, forecast-accuracy analysis, anomaly detection. LSTM, Prophet, ARIMA, SARIMAX, ETS, Croston, LightGBM, XGBoost, Random Forest, SVM, neural networks. TensorFlow, Keras, Scikit-learn, Statsmodels, AutoML.

FINTECH & QUANTITATIVE FINANCE

Risk modelling, wagering-behaviour analytics, closing line value, fundamental & technical market analysis, algorithmic trading (MQL5 Expert Advisors), working-capital decisions under uncertainty.

GENERATIVE AI & AGENTS

GPT-4 / GPT-4o, Azure OpenAI, Claude, RAG, LangChain, LangGraph, AutoGen, CrewAI, agentic architectures, Model Context Protocol (MCP) servers, embeddings, semantic search, Pinecone, FAISS, ChromaDB.

PROGRAMMING

Python, SQL, Rust, TypeScript, JavaScript, C#, .NET, C (bare-metal ARM), MQL5.

SYSTEMS & COMPILERS

Compiler construction (lexing, parsing, type checking, MIR, region inference, code generation), LLVM, FFI / `extern "C"`, Apache Arrow zero-copy interop, multithreaded data processing, PyO3.

DATA ENGINEERING

BigQuery, Databricks, PostgreSQL, MySQL, SQL Server, MongoDB, SAP data integration, ETL pipelines, data-quality tooling.

CLOUD & MLOPS

GCP, Vertex AI, Azure ML, Azure DevOps, MLflow, Docker, Kubernetes, AKS, AWS (EC2, S3, RDS, Lambda, IAM), FastAPI, CI/CD.

PROFESSIONAL EXPERIENCE

Founder — StockAI & WorkforceAI

Independent · Jun 2026 – Present · stockai.es · workforceai.es

- My main work since leaving Softtek: my own two products, built and run end to end — models, backend and interface.
- **StockAI** — inventory and purchasing SaaS for Latin American businesses running thousands of SKUs without a data team. Every day it sorts each product into “Pedir YA”, “Pedir pronto”, “OK” or “Sobrestock”, with multi-warehouse transfers, a scenario simulator, an AI analyst in plain Spanish and a daily email / WhatsApp summary.
- StockAI's forecasting engine: up to nine models compete per product (LightGBM, XGBoost, ARIMA, SARIMAX, Prophet, ETS, Croston, LSTM and a global catalogue model), chosen by demand pattern and backtested over the buying horizon with a stockout weighted 3× an overstock; the winner must beat two naive baselines. Stockout days are flagged so lost sales are not learned as low demand, and lead times are learned from real deliveries.
- **WorkforceAI** — a virtual office of AI employees for small teams: an LLM plans each one-sentence request as a dependency graph that the backend validates and runs, replanning a failed task on its own, and returns a report of what was done and what it cost.

- WorkforceAI's guardrails: tool calls are proposals the backend checks against the agent's tools, policy and autonomy level; cost is reserved before every model call against request, agent and organisation caps; irreversible actions wait for a human; outside text is masked and scored for prompt injection. Bring-your-own-key across DeepSeek, Claude, OpenAI, Kimi or any OpenAI-compatible endpoint. Go, FastAPI, CrewAI, Next.js, Postgres with row-level security. In early access.

ML / AI Consultant (part-time)

Photon Rush · Jul 2026 – Present · consulting, alongside StockAI & WorkforceAI

- Designing a multi-model system to detect “sharp” bettors — players with a sustained edge over the house — so a risk and trading team can act through limits, line adjustments and account review.
- Feature engineering on wagering behaviour — bet timing relative to line movement, stake concentration by sport/league/market, closing line value — with a SHAP-based explainability layer wired to an LLM so each flagged player arrives with a plain-language rationale.

Data Scientist

Softtek (Walmart vendor) · Mar 2026 – May 2026

- Built end-to-end demand forecasting for Walmart's agricultural supply chain — distribution-centre inventory, retail demand, store inventory, supplier constraints, lead times and raw-material availability — supporting procurement and replenishment at Hortifruti Coris (Cartago) and CEDI Coyol (Alajuela).
- Selected LSTM, ARIMA, ETS and Croston per series by seasonality and intermittency rather than by preference, with feature engineering, outlier detection, missing-value treatment and normalisation across large-scale operational data.
- Designed replenishment and purchasing recommendations by combining forecast output with inventory availability and operational constraints.
- Contributed to administrative shrinkage detection, a problem estimated near **$8M** in annual losses in Costa Rica.
- Deployed enterprise agentic AI on Model Context Protocol servers, letting agents query internal supply-chain databases directly for the Célula GenAI team, and integrated LLM-assisted forecast review.
- Worked across GCP, BigQuery, Vertex AI, SAP, Jupyter and Colab.

App Automation Senior Analyst · Machine Learning Engineer

Accenture · Aug 2024 – Feb 2026

- Architected end-to-end ML for reverse supply-chain optimisation across HP, Philips and Cisco: delivery-time forecasting, bottleneck detection, throughput optimisation and cost prediction.
- Built the pipelines around them — ingestion, feature engineering, training, tuning, deployment, monitoring and retraining — at enterprise scale.
- Designed and maintained production MLOps on Azure ML, MLflow, Azure DevOps, Docker, Kubernetes, AKS and CI/CD, serving models through FastAPI and Flask behind enterprise REST APIs.
- Developed the generative-AI side: GPT-4 and Azure OpenAI, Claude, LangChain, LangGraph, AutoGen, CrewAI and RAG, with vector retrieval on Pinecone, FAISS and ChromaDB.
- Designed agentic systems with tool calling, memory persistence, workflow orchestration, multi-agent collaboration and enterprise integrations.

Full-Stack Developer

TeamArt Outsourcing · Jan 2023 – Aug 2024

- Delivered enterprise web and mobile applications in Angular, React, React Native, RxJS, Redux and Material UI.
- Built microservice architectures in Node.js, TypeScript, AWS Lambda and PostgreSQL on Domain-Driven Design, with REST APIs in Python and Django using CQRS and event-driven patterns.
- Managed AWS infrastructure (EC2, S3, RDS, IAM, Lambda, EFS) and worked directly with construction and e-commerce clients through to production.

Freelance Full-Stack Developer

Fiverr, Upwork & Workana · Aug 2020 – Jan 2023

- Delivered full-stack solutions for international clients across several industries.
- Front ends in React, Angular, TypeScript, JavaScript, CSS/SCSS and Bootstrap; back ends in .NET Core, Node.js, Django and Python; databases in MySQL, PostgreSQL and SQL Server.
- Implemented JWT and Auth0 authentication and containerised deployments with Docker.

INDEPENDENT ENGINEERING & RESEARCH

uNOS — Universal Neural Operating System

`Creator · brain-computer interfaces`

- A hardware-agnostic layer between any neuroimaging device (EEG, ECoG, implant, fNIRS, EMG) and any effector (cursor, robotic arm, wheelchair, exoskeleton), behind one contract: raw signal in, structured intention out.
- Its least convenient finding is published as the headline: across 22 subjects and three public datasets *no decoder wins* — the gap between people (14–15 accuracy points) dwarfs the gap between algorithms (1.6 points), so the advantage is switching decoder per person, not picking a best one.
- Conformance-tested against its own drivers, three of four of which failed on the first run.

10.5 ms p50 end-to-end latency (p99 13.6 ms) · **0 / 600,000** samples lost over a 10-minute endurance run

v0.1 — proven on simulated devices and public recordings, not yet on real hardware or with an outside integrator.

FastDNA — a genomic k-mer counter in Rust

`Creator · Rust, PyO3, Apache Arrow`

- Reads raw FASTQ, counts k-mers in parallel across every CPU core, and hands the result to Python as a zero-copy Arrow table — no serialisation tax between the systems language and the notebook.
- Correctness is checked against KMC3 on real *E. coli* sequencing data, where the counts match exactly.
- One of its own published benchmarks was wrong: it was diagnosed, retracted publicly, fixed, and the retraction still sits in the repository beside the corrected number.

44–63× faster than a Python k-mer counter at 8 threads · **2.4×** faster after the retracted benchmark was fixed

Not yet on PyPI — build from source. Still behind dedicated counters such as KMC3 at the largest scale tested.

Science — a compiled language and its compiler (side project)

`Creator · Rust, LLVM · github.com/Angel-Zeledon/sciencec`

- Fourteen Rust crates taking a source file from lexer to native binary through LLVM; ownership without a garbage collector, with regions inferred rather than annotated.

2,173 tests passing · LLVM 18.1 backend

PUBLICATIONS & WRITING

Cauce: Topic-Delegable Liquid Democracy with Reputation-Weighted Delegates

`Sole author · 2026 · 19 pp., preprint`

- A liquid-democracy design where delegation is scoped per policy topic, overridable bill by bill, and — the central departure from prior systems — accompanied by a published, falsifiable, Brier-scored track record for every delegate.
- States the delegation-resolution algorithm formally and proves four propositions and a corollary: that an override is local, that the weight cap applies exactly and independently of ballot content, that path resolution always halts including on cycles, and that under a correlated blind spot no rule can be correct and every rule loses the same multiplicative factor.
- On synthetic populations, reputation-weighted delegate choice held **99.6%** correct outcomes where direct voting reached 0% and popularity-driven delegation 65.4%, at an unaffiliated-population competence of 0.35 — and the weight cap turned out to be largely orthogonal to that result.
- Validated by nine hand-built unit tests and a cross-validation of a vectorised implementation against a literal one to within 1e-6 over 300 bills, a check that earned its keep by catching a real bug.

De la realidad a la abstracción — a complete university mathematics course

`Author · self-published in Spanish · first edition, 2026 · not peer-reviewed`

- Written from scratch, from logic and set theory up to category theory. **2,080 pages, 38 chapters, 903 numbered theorems, 468 exercises.** Published alongside an audit of its own defects: about a third of its numbered statements are proven far from where they are used or not at all, and a companion volume now supplies the missing proofs.

EDUCATION & LANGUAGES

- B.Sc. Computer Engineering — Universidad Hispanoamericana *(final year)*
- Computer Engineering — Tecnológico de Costa Rica *(2 years completed)*
- Spanish — native
- English — professional working proficiency