

Manuel José Vargas Alvarez

AI Engineer · ML Engineer · Data Scientist

Buga, Colombia (Remote) | English, Spanish | manuel@manuelvargas.dev | +1 (805) 245-7835
manuelvargas.dev | linkedin.com/in/manuel-v-01091999 | github.com/ArepaMan

PROFESSIONAL SUMMARY

MS Data Science (University of Pennsylvania, GPA 3.93) with a portfolio of shipped, measured AI systems: a **52.9M-parameter** transformer trained from scratch with public repo, live demo and published weights; an anomaly detection pipeline over **509,358** public contracts; and a production GraphRAG platform evaluated on a hand-annotated corpus. Sole developer on two live multi-tenant SaaS products. GMP validation background (FDA 21 CFR) — leakage control, measured baselines, traceable evidence. Seeking AI, ML and Data Science roles (remote).

TECHNICAL SKILLS

ML & Data Science: Python, PyTorch, scikit-learn, pandas, Polars, DuckDB, transformer architecture, BPE tokenization, ablation studies, leakage detection, perplexity, ROC-AUC, calibration, anomaly scoring, gold-set benchmarks

AI & LLM: Retrieval Augmented Generation (RAG), GraphRAG, vector databases (Qdrant), knowledge graphs (FalkorDB), embeddings, hybrid retrieval, prompt engineering, agents, guardrails (NeMo), OpenAI API, Gemini, Qwen, n8n

Engineering: FastAPI, React, Next.js, TypeScript, SQL, Supabase, PostgreSQL, row-level security, Edge Functions, Docker, Cloudflare Workers, Streamlit, Hugging Face, CUDA, Git, CI/CD

PROJECTS

Sol — Transformer From Scratch

2026

PyTorch, CUDA, BPE, Streamlit | github.com/ArepaMan/sol | sol-52m.streamlit.app

- Built a **52.9M-parameter** decoder-only transformer end to end with no pretrained weights: validation perplexity **3.719** (95% CI 3.693–3.745) against a 23.4 trigram baseline on the same 15,141-document set.
- Caught cross-split leakage — **6,304** validation documents were exact duplicates of training documents — and dropped them before any reported metric.
- Measured a seed-variance floor (4.490 ± 0.0045 ppl over 3 identical runs) before interpreting ablations; learning rate clears it by $\sim 271\times$, data scale by only $\sim 18\times$. 40k iterations, 1.31B tokens, 16.9 h on one 8 GB laptop GPU; **154 tests** with a CI drift guard.

Elephant in the Room — Procurement Risk Detection

2026

Python, DuckDB, Polars, scikit-learn, SECOP II

- Pipeline over **509,358** Colombian health contracts (COP 59.8 trillion): 26,686 raw signals reduced to 48 findings through five false-positive gates measured against per-stratum base rates.
- Diagnosed gates suppressing real detections; a minimum-population rule and a corrected denominator raised post-adjudication recall from **60% to 100%** across 11 injected fraud patterns.
- Linear UNSPSC classifier on 492,025 declared labels — **90.8% accuracy** against a 53.3% majority class; the model measures, a deterministic rule with published thresholds decides. Adversarial verification runs against the published export, not the code; 176 tests, 96.7% join coverage.

Lante — Company Brain (GraphRAG)

2025 – 2026

Python, FastAPI, Qdrant, FalkorDB, Docker

- WhatsApp and PDF ingest into a tenant-scoped knowledge graph: entity **F1 0.933** (0.923–0.944 over 3 LLM runs) against a 0.692 heuristic baseline on a 93-row hand-annotated corpus, 48 rows deliberate negatives.
- Hybrid retrieval fusing Qdrant vectors with FalkorDB commitment-edge traversal, citing sender and timestamp and refusing when evidence is missing. Rebuilt a contaminated gold set behind a corpus hygiene test; **431** passing offline tests.

Bodega Boost — Retail OS

2025 – 2026

React, TypeScript, Supabase, Edge Functions | bodegaboost.com

- Live multi-sided LATAM platform serving six account types from one codebase: **163 Edge Functions, 698 SQL migrations**, offline-first POS, WhatsApp commerce, Creem billing.
- Row-level-security multi-tenant schema with service-role-only writes on entitlement tables and a drift test blocking self-granted paid features; 54 Playwright specs, 151 unit test files.

Iter AI Parking

2026

TanStack Start, Supabase, Cloudflare Workers, Gemini | iter-ai-parking.manoloxzzz99.workers.dev

- Shipped a complete four-portal product — driver, owner, staff, ops — with reservations, billing and offline mode in **under 40 hours** of total solo build time.
- Zero mapping-API spend via a Postgres bounding-box plus haversine RPC; Gemini copilot with NeMo guardrails; authorization review clean of exploitable PO/P1 findings in the persona pass.

PROFESSIONAL EXPERIENCE

Contract Validation Engineer — cGMP Consulting Inc.

Jan 2022 – Dec 2024

Illinois, USA | Clients: AbbVie, Fresenius Kabi, Pfanstiehl, Uberlube

- Executed IQ/OQ/PQ and Periodic Validation Reviews; authored FMEA and risk-based validation plans under FDA 21 CFR.
- Drafted SOPs, change controls and deviations; supported CAPA closure with cross-functional QA teams.

EDUCATION

University of Pennsylvania — MSE, Artificial Intelligence, Machine Learning and Data Science

In progress

Philadelphia, PA | GPA 3.93 / 4.00

University of Wisconsin–Madison — BS, Chemical Engineering

2021

Madison, WI | GPA 3.72 / 4.00