Perry Johnson

Principal AI Engineer

Missoula, MT perryrjohnson7@gmail.com perryrjohnson.com linkedin.com/in/perryrjohnson github.com/perryrjohnson

SUMMARY

AI engineer with 8+ years building production systems and the data infrastructure under them, mostly in physical industries that run on messy data: heavy machinery, agriculture, shipping and trade finance.

EXPERIENCE

Principal AI Engineer, Digital Iron 2025 – Present

- Building the whole AI system behind aftersales intelligence for heavy machinery, from ingesting service records to agentic querying and fleet analytics (Django, Celery, Postgres, Meilisearch).
- Designed an agentic query engine that answers questions across structured and unstructured data, citing source tickets.
- Built an offline evaluation bench for fine-tuning that scores each pipeline step and tests candidate models, prompts and settings against the current system before they ship.
- Shrank embedding vectors 6x with learned compression while improving retrieval quality.
- Built an algorithmic fleet intelligence engine that finds failure chains, repeat offenders and cost anomalies.
- Built a provider-agnostic LLM layer with per-model circuit breakers, error-aware fallback and response caching.
- Built the post-processing pipeline behind extraction (equipment inference, hallucination checks against source text, financial reconciliation) and made ingestion recover on its own from worker and provider failures.

Co-founder & ML Engineer, Fishtail 2021 – 2025

- Built the data and ML systems behind a trade finance platform for SMB shippers and forwarders.
- Wrote a document engine that turns unstructured trade documents into structured data, at first using a fine-tuned vision transformer, then swapped out for fine-tuned LLMs.
- Canonicalized company names across 100M+ bills of lading with a transformer fine-tuned on internally labeled data.
- Reconstructed voyages from raw vessel tracking data and built a global port health index for maritime supply chains.
- Automated trade finance repayments end to end with a reconciliation system.

Lead Data Scientist, Farm Dog (acquired by Deveron) 2019 – 2021

- Built a geospatial pipeline moving agronomists' field observations and weather from the production app into PostGIS.
- Scored pest and disease severity, tracked trends, flagged outbreaks near a farm's fields and built a map-based app.

Data Scientist & ML Engineer, Independent 2018 – Present

- Advised on LLM architecture: structured outputs, provider-agnostic design, evals and when not to let an LLM decide.
- Uncovered credit risk signals across years of transaction and payment data.
- Collaborated with a leading cancer radiologist on image segmentation for colorectal anatomy and lesion detection.
- Wrote a Python library that stress-tests hospital software by simulating patient flows across hospital layouts.

Quantitative Analyst, Strix Leviathan 2018

- Built and backtested algorithmic trading strategies at a crypto hedge fund.

SKILLS

Languages	Python, Go, TypeScript, SQL
LLM systems	OpenAI, Anthropic, Groq, Fireworks, fine-tuning, embeddings, RAG & agentic retrieval, evals, structured outputs, multimodal/VLM pipelines, provider fallbacks, retries, circuit breakers, caching
ML	PyTorch, scikit-learn, Hugging Face, vision transformers, document AI & OCR, entity resolution
Math & statistics	Inference, regression, hypothesis testing, experiment design, econometrics, linear algebra, optimization
Data & pipelines	ETL, Airflow, Dagster, Celery, SQS, PostgreSQL, PostGIS, Redis, Meilisearch, Weaviate, pgvector
Systems & infra	Distributed, fault-tolerant pipelines, Django, FastAPI, Docker, Kubernetes, Terraform, AWS, GCP

EDUCATION

University of Washington, Foster School of Business — M.S., Entrepreneurship · Innovation focus

Lake Forest College — B.A., Economics · H. Murray Herlihy Prize in Economics