

Erik Rice

Senior Applied AI Scientist | AI Analytics Engineer | Reliable AI Systems

Chicago, IL | [linkedin.com/in/erik-rice-97a295282](https://www.linkedin.com/in/erik-rice-97a295282) | erik.s.rice@gmail.com | 612-750-7621

Professional Summary

Applied AI scientist and AI analytics engineer building reliable agentic systems, evaluation frameworks, and predictive workflows in high-stakes environments. Combines Python-based AI engineering - orchestration, typed contracts, retrieval, structured outputs, and evaluation harnesses - with forecasting, statistical validation, and decision-focused analytics. Experience spans federal public-health modeling, frontier coding-agent evaluation, governance-sensitive AI workflows, and reliability-focused system design.

Core Technical Toolkit

Python | PyTorch | LangGraph | Pydantic | DSPy | AWS | Agentic AI Systems | LLM Evaluation | Evaluation Harnesses | Typed Data Contracts | Structured Outputs | Retrieval-Augmented Generation (RAG) | Tool Selection & Tool Use | AI Reliability & Observability | Forecasting & Predictive Modeling | Statistical Validation & Experimentation

Professional Experience

World Trade Center Health Program (Planned Systems International) | Data Scientist (AI & Predictive Modeling) | Remote | Nov 2024 - Present

- Built the program's first AWS-hosted forecasting architecture and a 36-model suite using PyTorch, Prophet, Auto-ARIMA, and AutoML, replacing manual cohort workflows with repeatable long-range projections for official Congressional reporting.
- Applied statistical validation and quantitative peer review to identify macro-growth overestimation, reconcile model discrepancies, and strengthen reporting defensibility.
- Developed a DSPy-assisted contract analysis workflow that extracts and structures spread-pricing evidence from unstructured vendor contracts and evaluates extraction quality against explicit criteria.
- Built a retrieval-augmented policy simulation workflow to compare financial liability scenarios under complex federal restructuring.

G2i | Senior AI Evaluator, Frontier Coding & Agentic Evaluation - Contract | Remote | May 2026 - July 2026

- Evaluated frontier coding and agentic models on multi-step software engineering tasks spanning repository analysis, code modification, debugging, tool use, and end-to-end task completion.
- Traced failures across planning, context retention, tool selection, implementation, testing, and recovery to identify the earliest decision that broke task completion.
- Audited prompts, tests, graders, expected outputs, code changes, and reasoning traces for leakage, brittle logic, false positives, construct-validity failures, and superficial benchmark passes.

- Synthesized recurring model failures and benchmark defects into structured feedback on capability, evaluation blind spots, and execution reliability under realistic engineering constraints.

Dominican University | Research Data Scientist / Student Success Data Analyst | River Forest, IL | May 2023 - May 2026

- Built a governance-first IRB review pipeline that chunks HRP-503 protocols, redacts PII with Presidio/spaCy, and routes cases through LangGraph for stateful, auditable risk review of HIPAA/FERPA-sensitive research protocols.
- Deployed Dominican's first predictive model—the Transition Readiness Model (TRM), a rule-based multi-source classifier with capacity-aware percentile recalibration—improving early-risk targeting by 42%.

BJB PROPERTIES LTD | Property Operations & Analytics Manager | Chicago, IL | Aug 2018 - May 2023

- Built Python/dbt and Yardi pipelines replacing manual reporting with automated performance trackers and lifecycle models.
- Analyzed occupancy trends, property budgets and operating costs, and maintenance history to mitigate turnover risk and support preventative maintenance planning.

Topshop | Inventory Operations Manager | Chicago, IL | Jun 2015 - Jul 2018

- Engineered custom tracking systems to analyze historical demand, reducing floor stockouts by 28% during peak volume.
- Used product demand, pricing, and fulfillment data to guide inventory planning and pricing strategy.

Selected AI Evaluation & Reliability Projects

- **Tool Choice Contract Trial:** Built a replayable Python/Pydantic tool-admissibility harness with typed task and tool-manifest schemas, policy-oracle separation, deterministic clause-level failure evidence, and canonical JSONL/Markdown artifacts.
- **Synthetic Goodhart Trap Simulator:** Developed a synthetic experiment that measures when proxy optimization improves the reported score while degrading the intended latent outcome, isolating metric pressure and construct-validity failure.
- **KPI Definition Change Trial:** Developed a metric-definition migration experiment that runs identical historical fixtures through old and revised KPI specifications, quantifying restatements and decision flips caused by definition drift.

Education & Credentials

- **DePaul University** | Bachelor's Degree, Psychology and Statistics | 2012
- **DataCamp** | Certified AI Engineer for Data Science (Highest Tier) | 2026