COLE HOFFMAN

cal.com/csh-1997-eng • linkedin.com/in/connect2colehoffman • github.com/csh-1997-eng • x.com/lordcolton_exe

Applied ML engineer specializing in quantitative investment systems, from research prototyping to production deployment. 4 years building ML/AI infrastructure at a $9T asset manager across fixed income, equity, and multi-asset workflows. Deliver the financial return on investment of a researcher with the organization and methodical precision of an engineer. Deep experience in prepayment/survival modeling, factor-based portfolio construction, demand forecasting, and ML interpretability at enterprise scale.

SELECTED IMPACT

- **20-model ensemble** (10 CatBoost + 10 Neural Nets) on TB-scale loan-level data, beating incumbent vendor (YieldBook) for MBS prepayment prediction. **$50M+ projected annual alpha opportunity.**
- **Demand forecasting system** predicting cash flows across **28K products and 36 brands** using 30+ years of Morningstar and FRED data. Directly supporting pricing decisions tied to multi-billion-dollar fund flows.
- **300× backtesting acceleration.** Distributed SageMaker pipelines running **360 models in parallel**, full backtest (1996–2024) in 4 days.
- **End-to-end systematic investment pipeline.** 2,000+ equities, 23 years of macro + fundamental data, factor-neutral portfolio construction, risk optimization. **Delivered in <14 days.**
- **Novel geospatial feature engineering.** Seq2vec + autoencoder embeddings from BLS, HUD, and IPUMS data, **applied to the geographic dimension of MBS prepayment research.**

EXPERIENCE

Senior Applied AI Engineer / Tech Lead   09/2024 – Present
Global Investment Systems, Vanguard

- **MBS Prepayment Modeling.** Firm paying for YieldBook licenses to model conditional prepayment rates. **Built** TB-scale data pipelines, ran experiments across CatBoost, XGBoost, and TensorFlow on loan-level and pool-level data. Developed custom **geospatial embeddings** (seq2vec + autoencoder on BLS/HUD/IPUMS) improving geographic feature representation. Used SHAP and probabilistic (kernel) SHAP for transformer explainability. **Delivered 20-model ensemble beating the incumbent in 18 months.**
- **GenAI Interpretability Layer.** Researchers losing days to manual experiment interpretation. **Built** an agentic system (AWS Bedrock) that ingests evaluation metrics, SHAP functional form plots (interaction, violin, bar), pipeline code, and researcher axioms. Orchestration layer cross-references outputs against axioms to validate hypotheses. **Produces executive-ready PDF reports. Evaluated with DeepEval and Arize.**
- **Systematic Investment Pipeline.** No shared research infrastructure for new hires. **Built** end-to-end in **<14 days**: daily pricing (2,000+ tickers, 2000–2025), Fama-French 5-factor returns, 143 macro features, GICS industry classifications. Models: linear (industry beta), XGBoost, autoencoder with macro factor head. Monthly expected return forecasts, factor-neutral portfolio construction via risk model + optimizer, full backtest and evaluation. **Delivered as a cloneable research platform.**
- **Hypothesis Review System.** Research team spending weeks on Toulmin diagrams by hand. **Built** a multi-agent system that extracts, critiques, and cross-checks hypotheses with researcher-in-the-loop feedback. **Adopted across 14-person research team.**
- **Tech lead** across 6 ML engineers and 2 researchers. Set engineering standards, guided architecture, mentored execution.
- **Consulted** as SME on 3 additional projects: syndication ML model, quantitative portfolio manager evaluation, and LLM-powered vulnerability analysis.

Applied AI Engineer   08/2023 – 09/2024
Global Investment Systems, Vanguard

- **Fund Cash Flow Forecasting.** Researcher had a custom TensorFlow/Keras model stuck in 5 notebooks with a 5-day training cycle. **Re-architected** into 2 SageMaker workflows: production inference (<24 hours) and distributed backtesting (**360 models in parallel, 1996–2024 in 4 days**). Rewrote processing in Polars. Built multi-dimensional SHAP analysis with aggregation across cluster, time, and firm dimensions. Predicting cash flows across **28K products and 36 brands.**
- **SEC 10-K RAG Platform. Built** end-to-end: EDGAR ingestion, SEC XML parsing, vector storage in Pinecone (indexed by company/year/quarter/section), LLM-powered retrieval with intelligent query decomposition. Handed off to MLE team for production scaling.

- **Productionized** 2 of Vanguard's 5 major ML solutions end-to-end: deployment pipelines, inference architecture, monitoring, reliability.
- **Coached** delivery of a tree-based churn prediction model linking web engagement telemetry to fund-flow momentum. **Integrated into weekly portfolio manager decision cycles.**

Full-Stack Engineer   12/2022 – 08/2023
Equity Investment Systems, Vanguard

- **Built** a full-stack benchmark analytics platform (TypeScript/React + Python) for 20+ portfolio managers to compare benchmarks against active portfolios and project future trades. **Projected $40M–80M annual impact.**
- **Designed** microservices for benchmark construction with event-driven architecture, observability, and production reliability.

Machine Learning Engineer   04/2022 – 11/2022
Enterprise Advice, Vanguard

- **Built** random forest models on client tax lot data for personalized portfolio recommendations. **Improved tax-loss harvesting by 90%.** Established early MLOps foundations.

Data Engineer   08/2021 – 04/2022
Chief Data Analytics Office, Vanguard

- **Built** enterprise ETL pipelines for compliance scanning and metadata cataloging. CCPA readiness.

R E S E A R C H   &   I N D E P E N D E N T   W O R K

- **Foundational Research Repositories. Reproduced** seminal ML and investment science papers in code: AlexNet, ResNet, LSTMs, Autoencoders, Transformers, CAPM, Fama-French 3/5, Black-Litterman, Markowitz Mean-Variance, Factor Zoo. Internal open-source repos used for team onboarding and experimentation.
- **Geospatial Embedding Model.** Seq2vec + autoencoder pipeline generating state/ZIP-level embeddings from BLS, HUD, and IPUMS data. **Applied to the MBS prepayment model's geographic feature engineering.**
- **Philadelphia Rental Price Prediction. Sourced** novel dataset combinations (Zillow, Amtrak, OpenData Philly, HUD, BLS, FRED) to build a custom model predicting rental price changes. Independent quantitative research.

E D U C A T I O N

Villanova University — M.S. Applied Statistics (1 year, while working full-time), 2022–2023
Pennsylvania State University — B.S. Cybersecurity & Risk Analysis, 2021

C E R T I F I C A T I O N S

Capital Markets and Security Analyst (CMSA) • AWS Cloud Practitioner

T E C H N I C A L   S K I L L S

Languages: Python, SQL, R, TypeScript/JavaScript, Go
Python Ecosystem: pandas, NumPy, FastAPI, Flask, Jupyter, pytest
ML/AI: TensorFlow, CatBoost, XGBoost, scikit-learn, SHAP, time series forecasting, survival/prepayment modeling, factor modeling, portfolio optimization, risk modeling
GenAI: RAG, agentic workflows, prompt engineering, fine-tuning, HuggingFace Transformers, LangChain, LlamaIndex, MLflow, Arize, DeepEval, Azure OpenAI, AWS Bedrock, Pinecone, Qdrant, ChromaDB
Cloud / Data / MLOps: AWS SageMaker, Lambda, ECS/Fargate, DynamoDB, SNS/SQS, Step Functions, Terraform, CI/CD, Spark, EMR, Glue, Athena, S3, Polars, Dask, Postgres
Engineering: Solution architecture, event-driven design, observability (CloudWatch, Splunk), domain-driven design, GIS (ArcGIS, GeoPandas), Agile/Scrum
Other: Supabase, Vercel, Next.js, Docker, Redis, Celery, Unreal Engine, Blender