

# COLE HOFFMAN

cal.com/csh-1997-eng • linkedin.com/in/connect2colehoffman • github.com/csh-1997-eng • x.com/lordcolton\_exe

Applied AI Engineer and technical lead on a high-velocity internal consulting team, specializing in rapid execution under ambiguity and owning solutions from discovery to prototype to live deployment. Operate as an embedded technical partner to researchers, portfolio managers, and executives: learn the domain, build the system, present the results, and drive adoption. Strong background across data engineering, machine learning, and generative AI. Experienced delivering cloud-native ML and GenAI platforms at enterprise scale through distributed training, orchestration, and MLOps. Known for translating stakeholder needs into demo-ready prototypes through tight feedback loops, and self-initiating systems that accelerate team velocity and tie directly to measurable business outcomes

## SELECTED IMPACT

- **\$50M+ annual value.** Led a 20-model ensemble that beat the incumbent vendor for MBS prepayment prediction. Presented results directly to the Fixed Income trading desk.
- **300x backtesting acceleration.** Took a researcher's 5-day, 5-notebook training cycle and rebuilt it into distributed pipelines running 360 models in parallel. Became the team's daily business point of contact.
- **Self-initiated GenAI interpretability system.** Built an agentic pipeline that transforms experiment outputs into executive-ready PDF reports comparing results against research hypotheses. Adopted by a 14-person research team.
- **Reputation-driven demand across 3 business units.** Brought in as subject matter expert on syndication modeling, portfolio manager evaluation, and InfoSec vulnerability analysis.

## EXPERIENCE

### Senior Applied AI Engineer / Tech Lead (Internal Consulting)

09/2024 – Present

*Global Investment Systems, Vanguard*

- **MBS Prepayment Model.** Firm was paying for vendor licenses (YieldBook) to model prepayment risk. **Partnered** with the trading desk and 2 researchers to replace it. Owned TB-scale pipelines, ran experiments across CatBoost and TensorFlow, **led** the geographic research dimension. **Delivered 20-model ensemble beating the incumbent in 18 months.**
- **Agentic Research Summarization.** Researchers losing days to manual experiment interpretation. **Built** an agentic system (AWS Bedrock) that processes metrics, SHAP plots, code, and researcher axioms into PDF reports with hypothesis validation. Includes model-vs-model comparison and interactive Q&A. **Adopted across the research team.**
- **Systematic Investment Pipeline.** No shared research infrastructure for new hires. **Built** end-to-end in **<14 days**: 2,000+ equities, 23 years of macro + fundamental data, expected return forecasting (linear, XGBoost, autoencoder), factor-neutral portfolio construction, risk optimization, and backtesting. **Delivered as a cloneable research platform.**
- **Hypothesis Review System.** Research team spending weeks on Toulmin diagrams by hand in PowerPoint. **Built** a multi-agent system that extracts, critiques, and cross-checks hypotheses with researcher-in-the-loop feedback at each stage. **Adopted across 14-person team.**
- **Tech lead** across 6 ML engineers and 2 researchers. Set standards, guided architecture, mentored execution.
- **Consulted** as SME on 3 additional projects across syndication modeling, PM oversight, and InfoSec.
- **Presented** regularly to executives on ML engineering talent strategy. Became a known internal advocate.

### Applied AI Engineer (Internal Consulting)

08/2023 – 09/2024

*Global Investment Systems, Vanguard*

- **Fund Demand Forecasting.** Took a researcher's 5-notebook, 5-day training cycle and **rebuilt** it into parallel SageMaker workflows (**360 models at once**). Backtest across 28 years in 4 days. Then **became the daily point of contact** with Fund Pricing & Economics, informing pricing strategy across 28K products and 36 brands.
- **SEC Document Explorer.** Talked to portfolio managers, **identified** the use case, **built** end-to-end: EDGAR ingestion, vector storage (Pinecone), LLM-powered retrieval by company/year/section. Handed off to my team to scale, then ran an internal road show **driving adoption across 5+ teams.**
- **Productionized** 2 of Vanguard's 5 major ML solutions end-to-end.
- **Coached** a junior engineer through delivery of a churn prediction model from web engagement telemetry. Guided architecture and solved key technical challenges. **Integrated into weekly PM decision cycles.**

- **ML Engineering Champion.** Regularly **pitched executives** on the value of ML engineers who understand investing, research, and engineering. Became a personal brand internally.

## Full-Stack Engineer

12/2022 – 08/2023

*Equity Investment Systems, Vanguard*

- **Benchmark Analytics Platform.** **Built** full-stack (TypeScript/React + Python) for 20+ portfolio managers. **Drove adoption** through recurring demos and iterative feature delivery. **Projected \$40M–80M annual impact.**
- **LLM Chatbot Prototype.** Asked to present on chatbot potential for Personal Indexing. Instead of a deck, **built and demoed a working chatbot** using HuggingFace transformer weights. Cemented my reputation for delivering beyond what's asked.
- **Designed** microservices for benchmark construction with event-driven architecture and production observability.

## Machine Learning Engineer

04/2022 – 11/2022

*Enterprise Advice, Vanguard*

- **Built** random forest models on client tax lot data. **Improved tax-loss harvesting by 90%.** Built 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.

## I N D E P E N D E N T P R O J E C T S

- **AI Poker Coach.** Designed, built, and **shipped** a Flutter mobile app end-to-end. Real-time GenAI feedback on poker decision-making.
- **Philly Rental Price Prediction.** **Sourced** novel datasets (Zillow, Amtrak, OpenData Philly, HUD, BLS, FRED) to predict rental price changes across Philadelphia neighborhoods.
- **Geospatial Embeddings.** Seq2vec + autoencoder generating ZIP-level embeddings from BLS/HUD/IPUMS data. **Later applied in the MBS prepayment model.**

## 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, TypeScript/JavaScript/Next.js, SQL, R, Go

**Python Ecosystem:** pandas, NumPy, FastAPI, Flask, Jupyter, pytest

**Product / Full-Stack:** Supabase, Vercel, Next.js, Flutter, Docker, Redis, Celery, Figma, Unreal Engine, Blender

**ML/AI:** TensorFlow, CatBoost, XGBoost, scikit-learn, SHAP, time series forecasting, survival/prepayment modeling, factor modeling, portfolio optimization

**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