Niles Liu

niles.liu@gmail.com | linkedin.com/in/niles-liu

Education

Harvard University, School of Engineering and Applied Sciences — Sep 2026 – Dec 2027
M.S. in Data Science (thesis track) — joint program of Computer Science and Statistics

Massachusetts Institute of Technology, GPA: 5.0 / 5.0 — Sep 2022 – May 2026
B.S. in Mathematics & Computer Science (18-C); Minors in Statistics & Data Science and Music

- **Mathematics & Statistics:** Real Analysis, Numerical Analysis, Seminar in Analysis, Differential Equations, Modern Mathematical Statistics, Probability & Random Variables, Inference, Combinatorial Analysis
- **Machine Learning & Computer Science:** Graduate Machine Learning, Deep Learning, Modeling with ML, Algorithms, Computability & Complexity Theory

Research Experience

Research Assistant, Learning & Adaptive Systems Group - ETH Zurich — June 2025 – Dec 2025
Advised by Prof. Ya-Ping Hsieh (with Vignesh Ram Somnath)

- Derived adjoint-matching equations on manifolds and extended Schrödinger bridges to time/state-dependent potentials, enabling gradient-based optimization for molecular conformation sampling
- Implemented manifold-based adjoint matching to reduce data-driven sampling bias and recover underrepresented modalities
- Integrated manifold-exploration techniques into DiffDock, improving sampling exploration in protein–ligand docking

Research Assistant, Institute for Medical Engineering & Science - MIT — Jan 2025 – May 2025
Advised by Dr. Li-Wei H. Lehman

- Researched uncertainty-aware methods for reducing false arrhythmia alarms in ICU monitoring, where calibration matters more than raw accuracy
- Applied deep-ensemble Fully Convolutional Networks and explored conformal prediction-based uncertainty measures to minimize false negatives

Work Experience

AI/ML Engineer, Starwood Property Trust – Miami, FL — June 2025 – Present

- Designed and led a learning-to-rank pipeline for real-estate comparables across millions of property records spanning several databases; built a multi-seed CV NDCG@10 evaluation framework after a noise-floor experiment showed prior single-seed ablations were chasing noise; promoted an 18-feature XGBoost ranker validated by paired sign and Wilcoxon tests
- Documented stochastic non-determinism in XGBoost's `tree_method='hist'` under multi-threading (Spearman $\rho = 0.56$ between fixed-seed runs); diagnosed cross-source normalization bugs across millions of property records during cross-database entity resolution
- Built a queue-based microservices system for PDF data extraction across rent rolls, financials, invoices, and IC memos with multi-LLM routing (Claude, OpenAI, Azure DI); shipped a multi-document classifier using an A/B ensemble of LLM zero-shot and Azure DI custom classification for bundled-PDF segmentation
- Designed a SQL Server property-similarity engine spanning two heterogeneous databases via a unified-view abstraction, with a configurable rule engine, FastAPI scoring API, and Flask UI for non-engineer rule editing

AI/ML Engineer, Burmester & Vogel – Cambridge, MA — Jan 2024 – Present

- Co-developed an Azure Durable Functions document-intelligence pipeline (receive → classify → extract → validate → callback) for 10+ maritime document types using Azure Document Intelligence and LLM-driven structured extraction; raised digitization coverage of the core document type to near-complete
- Owned an in-process TF-IDF + Logistic-Regression remarks classifier (28 classes, sigmoid-calibrated) that replaced per-remark LLM calls as the fast path with confidence-gated LLM fallback — cutting latency and cost while measuring *more* accurate than the LLM on the high-stakes class (recall 0.83 vs 0.46); built the offline benchmark + disagreement-adjudication harness that justified the rollout
- Ported the document-intelligence extractors (vessel-valuation and mortgage/facility) into a consolidated Python monorepo (uv workspaces, SQLAlchemy 2.0, Pydantic v2, Azure Functions)

Teaching Experience

Teaching Assistant, MIT Experimental Study Group — Sep 2023 – May 2026

- TA for General Chemistry, Linear Algebra (Head TA), and Multivariable Calculus; led weekly office hours for 50+ undergraduates and managed grading coordination as Head TA
- Received top student evaluations each semester; invited to independently teach an ES.1802 (Multivariable Calculus) section in Fall 2025

Skills & Activities

Languages & Tools: Python, Julia, SQL, LaTeX, Git, Docker; Azure (Functions, Document Intelligence, Blob, Key Vault)
ML/AI: PyTorch, TensorFlow, scikit-learn, XGBoost, NumPy, SciPy, Pandas, PydanticAI, Elasticsearch, SQLAlchemy 2.0
Methods: Generative Modeling & OT, Uncertainty Quantification (Conformal, Deep Ensembles), Learning-to-Rank, Multi-Agent LLM Systems, Statistical Inference
Activities: MIT Varsity Fencing (NCAA D-III squad leader, 2022–present); violin (MIT Chamber Music Society, Emerson/Harris Program); MIT Consulting Group (multiple leadership roles, 2023–2025)