

Jinu Hyun

(+1) 213-905-2425 | jh10305@nyu.edu | [LinkedIn](#) | [Portfolio](#)

EDUCATION

New York University

Expected: May 2027

Master of Science, Data Science

- Coursework: Intro to Data Science, Optimization and Computational Linear Algebra, Learning from Small Data, Machine Learning, Big Data, Data Engineering, Deep Learning for Computer Vision

The University of Texas at Austin

May 2025

Bachelor of Science, Mathematics | Minor: Computer Science

EXPERIENCE

Research Assistant | Razavian Lab, NYU

Mar 2026 – Present

- Developing **self-supervised pretraining** strategies for a 3D neuroimaging foundation model on 300k+ multi-sequence MRI scans, evaluating **masked prediction objectives** and **regularization** techniques across 13 clinical prediction tasks.

Research Assistant | RiskEcon Lab @ NYU Courant

Mar 2026 – Present

- Built an **entity-resolution pipeline** across multi-source tabular data (**fuzzy text similarity, typed constraints, ranked candidate scoring**) to link records at the right node for downstream modeling.
- Shipped **scalable matching + QA workflows** (auto-accept vs. human review) and **cleaned analytical datasets** that improved coverage and traceability for pricing analytics.

Data Science Intern | EV&Solution

Jun 2025 – Aug 2025

- Automated exploratory analysis of thousands of logistics datasets via a **reusable pipeline**, identifying critical anomalies.
- Applied **DBSCAN clustering** to reduce GPS noise by ~98% per trip, enabling extraction of vehicle parking coordinates.
- Created a **PostGIS-optimized geospatial database** for low-latency proximity searches and address-lookups.
- Used a **Flask backend** to generate real-time parking recommendations into a legacy map interface.

Research Intern | Human-Data Interaction Lab, Yonsei University

May 2025 – Aug 2025

- Built **evaluation framework** for **Vision Transformers** to quantify visual shortcut reliance on chart interpretation tasks, augmenting the dataset and utilizing **PyTorch Grad-CAM** for analysis.

Business Development Intern | QANDA (Mathpresso)

Sep 2024 – Dec 2024

- Fine-tuned **GPT-2** to generate context-aware responses, reducing manual overhead for marketing content.
- Integrated **Sentence-BERT** relevancy ranking to identify target conversations and auto-generate promotional responses.
- Analyzed user data with **SQL** in **BigQuery** to design segmented marketing campaigns, boosting email CTR by ~10%.

Sergeant | Republic of Korea Army

Aug 2020 – Feb 2022

- Member of 6th Corps 28th Division Artillery Brigade

PROJECTS

Earnings Research RAG Agent

Feb 2026

- Engineered a **multi-agent RAG system** over SEC filings and financial news to answer natural language queries about earnings surprises and stock movements.
- Built a **RAG pipeline** with **XBRL-aware sliding window chunking** to preserve table structure in financial filings, with metadata filtering and a **cross-encoder reranker** to improve retrieval precision.

Zero-Shot Malware Image Classification

Oct 2025 – Nov 2025

- Adapted the **Dense Attribute-Based Attention** framework for zero-shot malware classification, achieving 59% accuracy comparable to the original benchmark.
- Manually constructed a **12-attribute semantic matrix** for 10 malware families as ground-truth supervision, revealing that spatial attention-based semantic-visual alignment breaks down on non-interpretable image domains.

NYC Transportation Pipeline

Jan 2025 – Apr 2025

- Built an **ELT pipeline** on **GCP** integrating millions of records, leveraging **Vertex AI embeddings** to solve data-unification challenges (fuzzy matching) and **dbt** for automated modeling.

TECHNICAL SKILLS

Languages & Databases: Python, SQL, PostgreSQL, MongoDB

Libraries & Frameworks: Pandas, NumPy, Scikit-learn, Matplotlib, SciPy, Hadoop, Spark, PyTorch, LangChain, LangGraph

Infrastructure & Tools: Google Cloud Platform, Git/GitHub, Docker