

Mario Hernandez

787-519-1049 | mario.hernandez4@upr.edu | github.com/pollyjuice74
linkedin.com/in/mario-hernandez-b0296b250

DISTRIBUTED SYSTEMS & REAL-TIME INFRASTRUCTURE ENGINEER

EXPERIENCE

ML Systems Engineer — Market Intelligence

Independent Research Project — Jan. 2024 – Present

- Built a real-time trading research system (50k+ LOC) on AWS for multi-exchange ingestion and GPU-based short-horizon forecasting.
- Architected S3-synchronized worker processes for fault-tolerant shared state, deterministic recovery, and scalable multi-exchange execution pipelines.
- Developed a deterministic replay engine with strict ordering guarantees for debugging, microstructure analysis, and latency modeling under partial observability.

NASA Space Grant Research Fellow

University of Puerto Rico — Nov. 2023 – May 2024

- Designed a transformer-based LDPC decoder for 5G NR using NVIDIA Sionna with linear-time attention.
- Implemented end-to-end training and evaluation pipelines using Tanner graphs, LLRs, and noisy-channel models.
- Published first-author paper **5G LDPC Linear Transformer for Channel Decoding** produced under a NASA Space Grant award.

Software-Defined Radio Research Intern

University of Nevada, Las Vegas — May 2025 – Aug. 2025

- Engineered and integrated PHY/MAC components into a modular SDR architecture for multi-agent LiDAR sharing over IEEE 802.11p.
- Implemented OFDM synchronization, packetization, and real-time point-cloud transmission.
- First-author of **A Software-Defined Radio Testbed for Distributed LiDAR Point Cloud Sharing**, designing and integrating SDR transceivers with a ROS/Docker backend for cooperative perception.

Computational Biology ML Research Intern

National Science Foundation — Sep. 2024 – Jan. 2025

- Developed ML systems for bee identification and OCR-based ecological monitoring.
- Optimized YOLO-based detectors using **TensorRT quantization** and deployed distributed inference on Google Cloud.
- Built a semi-automated labeling pipeline integrating CVAT, pre-trained models, and human-in-the-loop feedback.

NSF Research Intern — ML & Signal Processing

National Science Foundation — Jun. 2024 – Jul. 2024

- Implemented deep-learning decoding prototypes for 5G LDPC channel models.
- Contributed to a HuggingFace transformer for bird-call classification achieving **96% accuracy** and **93% precision**.

TECHNICAL SKILLS

Languages: Python, C/C++, Rust, Go, Solidity, SQL

Systems & Cloud: Linux, Docker, Kubernetes, CUDA, JAX, DeepSpeed, ONNX, MLflow, Spark, AWS, GCP

ML/AI: PyTorch, TensorFlow, Sionna, TensorRT, Scikit-learn, XGBoost, LightGBM

Web3 & Backend: Smart contracts, Web3.py/Ethers.js, DeFi protocols, on-chain analytics, Flask, Django, REST APIs

EDUCATION

B.S. in Computer Science and Mathematics | GPA: 3.58

University of Puerto Rico | 2022 – 2026

Relevant Coursework: Stanford CS231n (Deep Learning for Vision), MIT OCW Finance, Harvard CS50 AI / Python / Web

PUBLICATIONS

Hernandez, M., Piñero, F.

5G LDPC Linear Transformer for Channel Decoding

arXiv preprint arXiv:2501.14102, Jan. 2025

Hernandez, M., Bryce, E., Stubberud, P., Saberini, E., Morris, B.

A Software-Defined Radio Testbed for Distributed LiDAR Point Cloud Sharing with IEEE 802.11p in V2V Networks

arXiv preprint arXiv:2509.14523, Sep. 2025

PROJECTS

Local GPU server build | Unix, CUDA, CuDNN, Nvidia GPU

- High performance compute cluster build for training NNs with local RTX 4070 GPU.

Distributed File System | Python, NoSQL

- Implemented a DFS with a metadata node server, data storage nodes, list and copy clients.

Deterministic Data Normalization Engine | Python, Pandas

- Implemented reproducible multi-asset normalization and timestamp-alignment pipelines for the real-time trading research system.

Wikipedia Search Engine | Python, Scrapy, Django, NoSQL

- Built a recursive web crawler and indexing system for Wikipedia pages using TF-IDF, Page-Rank scoring and a NoSQL backend, implementing search, ranking, and structured data retrieval.