

Pablo Leyva

908-525-6062 | pl33@njit.edu | linkedin.com/in/pablo-leyva | github.com/pleyva2004 | pabloleyva.io

EDUCATION

New Jersey Institute of Technology

B.S. in Applied Mathematics (Statistics) & Computer Science (AI Concentration)

Newark, NJ

Aug 2023 – May 2027

Relevant Coursework: Data Structures & Algorithms, Computer Vision, Linear Algebra, Probability & Statistics, Database Systems, Software Engineering, Calculus III, Numerical Methods, Objected Oriented Programming, Deep Learning PhD Coursework

TECHNICAL SKILLS

Languages : Python, Typescript, JavaScript, Java, C++

Frameworks : React, NEXT.JS , Node, PyTorch, TensorFlow, scikit-learn, LangChain, LangGraph, MCP

Libraries : Jupyter Notebook, Pandas, NumPy, Pytest, scipy, FastAPI, Flask

Machine Learning : Linear Regression, Logistic Regression, K-Means Clustering, K-Nearest Neighbors (KNN)

AI/ML : Natural Language Processing (NLP), Neural Networks (DNN, CNN, RNN), Deep Learning, Reinforcement Learning

MLOps : Model Deployment, Data Pipelines, Docker, CI/CD, AWS, GCP, Azure, GitHub Actions, linux, kubernetes, unix

EXPERIENCE

Undergrad Researcher (CAHSI) - Multilingual LLM Training and Bias Analysis

October 2025 – Present

Research Candidate

Newark, NJ

- Build and train a **foundational multilingual model** on Spanish, Portuguese, Italian, Farsi, and Arabic, with the primary objective of enabling the model to "think" and reason in native languages rather than through translation
- Compare and analyze **bias patterns** between models trained exclusively in English versus multilingual models trained on diverse language corpora to evaluate cultural and linguistic bias disparities
- Train specialized **Small Language Models (SLMs)** on individual target languages and benchmark performance against a foundational multilingual LLM to **compare efficiency** versus **accuracy tradeoffs**
- Apply **research techniques** and **algorithm design** to evaluate model performance across multilingual tasks, measuring both computational efficiency and linguistic fidelity in native language reasoning

Undergrad Research and Innovation (URI) - Efficient SLM for Agentic Workflows

Sep 2025 – Present

Research Candidate

Newark, NJ

- Apply **research techniques** and **algorithm** design to develop an 3B parameter SLM architecture
- Integrate lightweight adapter layers (e.g., LoRA) to specialize the model for structured business workflows
- Evaluate the agentic reliability of the model in executing long, deterministic step-by-step tasks vs. larger LLM baselines
- Benchmark on-device performance to **improve performance** and demonstrate feasibility on consumer hardware

Apple

AI/ML Product Engineering Intern

May 2025 – August 2025

Cupertino, CA

- Combined **Large Language Model(LLM)s** with **Model Context Protocol (MCP)** using Typescript and Python
- Used **LangChain & LangGraph** to facilitate agent infrastructure and validation
- Led the design and evaluation of a Graph-RAG architecture to enhance knowledge discovery and information relevance across partner use cases, focused on **Partner Facing RAG ChatBot** to improve business insights and decision making

PROJECTS

Princeton Hackathon - NeuroCache | WebGPU, Transformers.js, WebAssembly, PyTorch, CoreML / MLX | Python

- Implemented **custom KV-cache manipulation layer** enabling selective memory retention/eviction, dynamic context compression via attention-based importance scoring, and memory slot management for persistent information storage across conversation sessions

GroupGPT – Collaborative Ideation Platform | Next.js, Supabase, Socket.IO, OpenAI API | Python, JavaScript

- Built and prototyped a collaborative chat interface with Next.js and Supabase Realtime, allowing multiple users to contribute simultaneously to a persistent conversation thread

Emotion-Aware Music Recommendation System | PyTorch, BERT, librosa, Spotify API, Flask | Python

- Architected **multi-modal deep learning pipeline** with separate CNN branches for mel-spectrogram analysis, BERT transformers for lyrics sentiment processing, and embedding layers for metadata fusion

COMMUNITY OUTREACH

President, Vice President - External Affairs, Event Coordinator

October 2023 – Present

Society of Hispanic Professional Engineers (SHPE)

Newark, NJ

- Confidently leading a team of 20 to organize professional events for 300+ students across campus using articulate communication
- Grew our External Company connections by 6x in only two months through confident outreach and relationship building
- Successfully breaking attendance records on a weekly basis by averaging 200% growth in member engagement