

Pablo Leyva

908-525-6062 | pl33@njit.edu | [linkedin.com/in/pablo-leyva](https://www.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 Research and Innovation (URI) - Beyond Next-Token Prediction

Sep 2025 – Present

Research Candidate

Newark, NJ

- Developing a novel LLM paradigm that reframes training from next-token prediction toward intent-inference, enabling models to generate clarifying questions as intermediate reasoning steps.
- Engineering transformer-based modifications and evaluation pipelines to support interactive, question-driven architectures, with metrics focused on reducing hallucinations and improving task accuracy.

Apple

May 2025 – August 2025

AIML Intern, Software Engineer Intern

Cupertino, CA

- Led a team of 3 interns to build MVP for Agentic Payment flow experience for higher leadership at Apple Pay
- Prototyped agent workflows combining **Large Language Model(LLM)-based product** recommendations with Apple Pay checkout experiences, leveraging Typescript to implement Model Context Protocol (MCP) for merchant catalog parsing and transactional automation
- 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**

Caterpillar Inc.

May 2024 – August 2024

Software Engineering Intern

Remote

- Analyzed the efficiency of software development within the company by utilizing Generative artificial intelligence(AI) to assess the impact and usability of code commits
- Optimized the Software Development Life Cycle (SDLC) by implementing data visualization dashboards that tracked code quality metrics, sprint velocity, and deployment frequency, applying Agile methodologies and DevOps best practices to improve team productivity and code maintainability

PROJECTS

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
- Integrated OpenAI's GPT models with conversation memory to provide context-aware summarization, clustering of ideas, and automated agenda generation
- Explored agentic design patterns to support role-specific AI assistants (e.g., Sponsorship Agent, Logistics Agent), showcasing use cases in **real-time conversational AI, context-aware agent workflows, and deploying Large-driven features in user-facing apps**

NYU Hackathon (1st Place) - QuSotch | *Numpy, scikit-learn, Pandas, Quantum Circuits* | **Python**

- Explored alternative quantum algorithms (Quantum Walks, Grover's Search) to further **optimize computational performance** for financial modeling
- **Reduced** computational **complexity** by roughly **75%** compared to classical Monte Carlo methods by implementing Quantum Monte Carlo (QMC) with parallelized stochastic modeling

Princeton Hackathon - Illume | *Python, Google Gemini, Pytest, Docker, Kubernetes, Poetry* | **Python**

- Implemented unit and integration tests with Pytest to validate RAG pipelines, data ingestion, and multi-modal input handling
- Documented AI modules, **API endpoints**, and deployment workflows to ensure smooth on boarding for contributors and maintain scalability
- Led deployment using **Docker and Kubernetes**, ensuring a reproducible and stable environment for multi-service AI applications

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
- Implemented **late-fusion neural architecture** combining audio, text, and metadata features through fully-connected layers, achieving emotion/genre classification for sequential recommendation modeling using Transformer encoders
- Deployed end-to-end system with **Seq2Seq prediction model** for next-song recommendations, integrated with Spotify, Last.fm, and Genius APIs, featuring real-time mood trajectory visualization and personalized playlist generation

Business - Live Translation | *bash, Docker, NVIDIA NIM* | **Python**

- Utilizing Open AI's faster-whisper Large model to perform live speech transcription with **84% accuracy**
- Implementing **multi-threading** to run the transcription and translation in parallel
- Running transcription on CPU and Translations and TTS on GPU to optimize and accelerate inference

Co-Founder
Association of Latino Professionals For America (ALPFA)

- Collaborated to start the first professional latino organization on campus for non-engineering students
- Started the external partners affairs network with local ALPFA chapters and companies

Sep 2024 – Present
Newark, NJ

President
Society of Hispanic Professional Engineers (SHPE)

- Built CLARA - AI Assistant to help with organization operations including email management, event planning, outreach, marketing and everything in between
- Leading team of 20 to organize professional development events for a memberbase of 300 + across campus
- 2x the number of students getting interviews and internships for the latino community
- Raised \$70,000 capital for the organization to take 100 students to the national conference and operate independently of the university

March 2025 – Present
Newark, NJ

Vice President - External Affairs
Society of Hispanic Professional Engineers (SHPE)

- Grew our External Organizations and Company connection by 6x in only two months
- Currently leading and managing a team of 13 to organize 60+ events and social events a semester to bring social and professional enrichment to our member base of 300+ students
- Successfully breaking attendance records on a weekly basis by averaging 200% growth in member engagement
- Started Webmaster position for NJIT SHPE yielding 81% growth in CS student participation
- Started SHPETina position for NJIT SHPE yielding 43% growth in Latina student participation
- Initiated mentorship program where up to 20+ students learned from the experience of the Executive Board
- Established relationships with exterior clubs from local universities to unite the Latino community in NJ/NYC area

May 2024 – March 2025
Newark, NJ

Event Coordinator
Society of Hispanic Professional Engineers (SHPE)

- Organize commodities such as food and workshops resulting in an average of 65 + student involvement
- Collaboratively work with a team of 12 to successfully leading SHPE to be the largest student organization on campus
- Run weekly events yielding NJIT SHPE to receive many accolades and be recognized by the Governor of Newark

Oct 2024 – May 2024
Newark, NJ