BRYAN A PINEDA.

832-909-8584 | pineda.bamp@gmail.com | https://www.bamp8.com

EDUCATION

Stanford University
Expected Graduation: June 2026
B.S. in Computer Science, Artificial Intelligence Track

EXPERIENCE

Alterion

L3 Software Engineer, Jan 2026 – Present

- Architected the core infrastructure and stateful memory systems for an autonomous AI agent, engineering dynamic task routing, tool calling, and long/short-term semantic retrieval.
- Engineered a multi-tier violation detection pipeline using z-score anomaly detection and deployed a large-scale guard model via vLLM across Google Cloud Run and Kubernetes.
- Built a comprehensive GitHub Actions CI pipeline with intelligent layer caching, automating vulnerability scanning on PRs and reducing average image build time by 60% (15m to 6m).
- Architected a multi-tier, role-based violation detection pipeline to secure LLM interactions, utilizing fine-tuned models and z-score anomaly detection on embedding drift.
- Executed end-to-end load testing suites using k6 to identify latency bottlenecks and ensure production stability under high-concurrency traffic.

Software Engineer Intern, Sep 2025 – Jan 2026

- Engineered an intelligent Slack bot using LangChain and a Gemini LLM to provide personalized restaurant recommendations for the team, serving as a primary AI agent for testing the company's core control panel product.
- Built the bot's custom backend by aggregating publicly available data on local restaurant menus and prices and integrated an LLM to learn and query teammate food preferences.
- Developed and completed V1 of an application-wide audit log system to track user actions, timestamps, and locations, enhancing platform security and traceability.

Projects & Research

Query-Aware Multi-Modal GNN Recommendation System
***Machine Learning Project,* Oct 2025 – December 2025**

- Architecting a recommendation engine that fuses RAG principles with Graph Neural Networks to retrieve movies based on conversational natural language queries.
- Constructed a unified heterogenous graph by merging MovieLens, INSPIRED2, and ReDial datasets, utilizing LLMs to generate embeddings, and FAISS for efficient candidate generation.

End-To-End Trajectory Prediction for Autonomous Driving
Computer Vision,* Apr 2025 – *Jul 2025

- Developed a Transformer-based model for predicting 5-second future trajectory of an autonomous vehicle using both the Waymo and nuScenes data. Extended the TransFuser architecture with BEV prediction and sequential outputs. Have achieved stable convergence with low prediction error on real-world driving scenarios.

SKILLS

- **Languages:** Python, C++, C, SQL, JavaScript, TypeScript, Swift.
- **Frameworks & Libraries:** PyTorch, TensorFlow, React, Node.js, Spark, SwiftUI, Pandas, NumPy.
- **Tools and Platforms:** Git, Docker, Kubernetes, K6, AWS, Linux, Google Cloud Platform.