

Juan Pablo Triana Martinez

Applied AI Data Scientist | Healthcare & Biomedical AI | Python | AWS & MLOps
Clovis, CA | USA | 93611 | Mobile: (+1) 714-443-1831 | jp triana.martinez@gmail.com | [Linkedin](#) | [GitHub](#) | [Portfolio](#)

PROFESSIONAL SUMMARY

Applied AI Data Scientist with 3+ years of experience executing R&D and ML cycle — data preparation, model development, evaluation, and cloud deployment— across engineering, biomedical, and applied AI domains. Strong Python and AWS foundation with hands-on clinical & biometrics dataset experience (ECGs and Phonocardiograms time-series, 100,000+ samples), production MLOps pipelines on SageMaker, and graduate-level training in NLP, Generative AI, MLOps, and Neural Networks (MS Applied AI, GPA: 3.89).

TECHNICAL SKILLS

Languages: Python (expert), SQL, R (working knowledge), MATLAB, Bash
ML & AI: PyTorch, TensorFlow, Keras, Scikit-learn, Hugging Face, LangChain, CrewAI, LangGraph, Ollama, XGBoost, Statsmodels
Data Science: Pandas, NumPy, SciPy, Matplotlib, Seaborn, Jupyter, Experimental Design, Model Evaluation (F1, AUC, IoU, Precision).
MLOps & Cloud (AWS): AWS SageMaker, S3, RDS, Athena, CloudWatch, GCP, GitLab CI/CD, Conda
NLP & Generative AI: Claude Code, OpenAI GPT Models, Ollama, LLaMA-2 70B (in-context prompting, data generation), FAISS, BERT, Bidirectional LSTMs, Transformers, Attention Mechanisms, VAEs, GANs, Diffusion Models (coursework)
Computer Vision: CNNs (ResNet, LinkNet, VGG, DenseNet), OpenCV, Semantic Segmentation, VAEs, GANs
Databases & Data Engineering: PostgreSQL, MySQL, SQLAlchemy, AWS RDS, AWS Athena, large-scale pipeline design
Project Tools: JIRA, Confluence, Agile/Scrum, Asana

RELEVANT WORK EXPERIENCE

ECG Colombia **Cali, Valle del Cauca, Colombia**
Junior Data Scientist **10/2025 – 03/2026**

- Developed a deep learning computer vision inference pipeline in PyTorch (LinkNet-ResNet) for automated extraction of structured information from scanned financial documents — directly transferable to clinical document AI.

Ohmic Technologies Inc. **Montreal, QC, Canada**
R&D Engineer, Biomedical Signal Specialist **01/2023 – 06/2024**

- Developed and maintained ML pipelines on clinical-grade physiological datasets (~100,000 samples), covering data preparation, feature engineering, model training, and evaluation — across the same model lifecycle phases required in healthcare AI.
- Designed an SVM binary classifier (Scikit-learn) on Fourier-transformed biosignal data achieving 88% test accuracy; documented evaluation methodology, failure modes, and proposed improvements.
- Built a Python data drift detection package using Euclidean, Cosine, Wasserstein, and Mahalanobis distance metrics to monitor deployed model consistency over time — a production post-deployment monitoring solution for real-world data environments.
- Engineered a predictive heart rate pipeline using autocorrelation and time-series analysis at 0.15-second temporal resolution from physiological sensor data, aligned with the kind of time-series clinical AI used in patient monitoring systems.
- Extracted and analyzed HRV metrics (time-domain, frequency-domain, non-linear features) from 100,000+ samples using NeuroKit2 — modeling physiological biomarkers from healthcare sensor datasets.
- Developed an automated reporting tool in Python (PySide6, Matplotlib, PDF export) with kernel density estimation; reduced analyst review time by 30%.
- Led structured clinical data collection campaign (5 participants, 5,000+ audio samples) under controlled protocols. Conducted cross-functional code reviews with QA and implementation teams, improving Python code quality (Pylint score) by 35%.

Tandem Launch Inc. **Montreal, QC, Canada**
R&D Engineer – Biomedical Signal Specialist **02/2022 – 01/2023**

- Designed and deployed a Python-based cloud data pipeline integrating **AWS S3, AWS RDS, and SQLAlchemy** for large-scale biomedical signal data management (200,000+ samples) — scalable healthcare data infrastructure experience directly applicable to EHR-scale data environments.
- Established two large-scale clinical data collection campaigns (150,000+ samples over 1.5 years), managing the full data lifecycle from acquisition protocol through cloud ingestion and storage.
- Implemented a Gaussian Mixture Model + cosine distance quality assurance system for automated validation of incoming sensor data across a 200,000+ sample corpus, ensuring pre-training data reliability.
- Migrated 10–20 core digital signal processing algorithms from MATLAB/OCTAVE to Python within an Agile sprint framework; maintained numerical accuracy and added unit test coverage.
- Developed a Python tooling application (PySide6, NumPy, Wavio) for clinical hardware signal generation; became the primary tool for chirp signal acquisition across engineering workflows.

Mitra Biotechnologies.	Guelph, ON, Canada
Biomedical Engineer Internship	07/2021 – 04/2022
- Authored 10 technical reports on IVD regulations, utility patents, and biomedical market analysis across Canada and the USA. - Produced 20+ pages of detailed FDA regulatory documentation on ELISA-based requirements for patent compliance — demonstrating awareness of healthcare regulatory frameworks relevant to clinical AI governance.	

ACADEMIC PROJECTS

End-to-End MLOps Pipeline — AWS SageMaker [GitHub] [Project]	01/2025 – 03/2025
Built a full model lifecycle pipeline on AWS SageMaker: LSTM training, batch inference jobs, real-time endpoint deployment, Athena experiment logging, and CloudWatch monitoring — covering intake through post-deployment management.	
IoT Elderly Fall Detection — Clinical Safety AI System [GitHub] [Project]	12/2024 – 02/2025
Co-developed a patient safety AI system using multi-camera spatiotemporal data; achieved Recall: 0.94 for fall prediction — safety-critical clinical monitoring directly relevant to patient safety use cases in health systems.	
Fraud Detection — Multi-Model Evaluation Study [GitHub] [Project]	10/2024 – 12/2024
Benchmarked Logistic Regression, SVM, Decision Tree, and LSTM classifiers (F1 up to 0.9988); rigorous overfitting and generalization analysis with documented recommendations.	
Emotional Segmentation — LLM + Bidirectional LSTM [GitHub] [Project]	01/2024 – 03/2024
Hybrid NLP pipeline using LLaMA-2 70B for synthetic data generation and a 4-layer BiLSTM for inference (IoU: 0.92, Precision: 0.94) — LLM-assisted clinical NLP analog.	
Synthetic Chest X-ray Generation — VAEs vs. GANs [GitHub] [Project]	04/2024 – 06/2024
Generative model comparison (VAEs, GANs) on CXR14 medical imaging dataset using VGG/DenseNet; VAE classifiers outperformed CNNs by 10–15% — clinical imaging AI research.	
Multi-Agent Financial Analysis — CrewAI [GitHub] [Project]	
Multi-agent LLM orchestration system demonstrating agentic AI workflow design.	

EDUCATION

UNIVERSITY OF SAN DIEGO	San Diego, CA, USA
Master of Science in Applied Artificial Intelligence (GPA: 3.89)	07/2024 – 04/2026
Coursework: NLP and GenAI, Data Analytics for IoT, MLOPs, Neural Networks & Deep Learning.	
STANFORD UNIVERSITY	Palo Alto, CA, USA
Postgraduate Certificate in AI/ML (GPA: 3.00), Credential Link	06/2023 – 06/2024
Coursework: Machine Learning, NLP, Computer Vision, Deep Learning, Generative AI Models.	
UNIVERSITY OF GUELPH	Guelph, ON, Canada
Bachelor of Engineering in Biomedical Engineering	09/2017–05/2021
Coursework: Bioinstrumentation Design, Systems and Control Theory, Signal Processing, Biomedical Signals	

SELECTED CERTIFICATIONS:

AWS Certified AI Practitioner Credential	09/2026
Artificial Intelligence Graduate Certificate — Stanford Online Credential	06/2024
Machine Learning Specialization DeepLearning.AI / Stanford CPD Credential	05/2023
Advanced Learning Algorithms DeepLearning.AI / Stanford CPD Credential	05/2023
AWS MySQL with S3, RDS Coursera Credential	02/2023

VOLUNTEER ACTIVITIES:

Fresno State University	Fresno, CA, USA
Applied AI Researcher/ Volunteer	01/2026 – Today
- Built a Data Pipeline in PyTorch to extract the Fourier Transform, Spectrogram, Mel-Spectrogram, Mel Ceptrum Coefficients, and Linear Ceptrum Coefficients to data augment 1D EEG Raw Voltage Channel Data, into 2D feature spaces. - Aiding in support of Python coding and research of Deep Learning Architectures, for EEG Data research paper applications.	

ADDITIONAL

- Bilingual: Native Spanish, English C2; Basic French - Volunteer: Giving Children Hope, Buena Park, CA — food distribution serving ~300 families/week (10/2024–12/2024)	
---	--