

Jose Esparza

Location: Trujillo, La Libertad – Peru

E-mail: joseesparzadc@gmail.com

ORCID iD: 0000-0002-9372-3763

Phone: +51 997 189 179

GitHub: [pebeto](#)

LinkedIn: [josesparza](#)

Senior ML Engineer | LLM / RAG Systems | Open-source Julia

Experience

- **Senior Machine Learning Engineer @ Maximo** February 2026 — Present
Architected and shipped a production agentic banking assistant for WhatsApp customer support: FastAPI + LangGraph agent, RAG over a Postgres + pgvector knowledge base, AWS-backed OTP delivery (SES/SNS), audio transcription via faster-whisper, and an LLM-generated monthly financial-report subsystem on EventBridge cron. Built an MCP (Model Context Protocol) server with FastMCP exposing internal tools to LLM agents; designed Metabase BI dashboards modeling fintech revenue (interchange, GMV, spreads) and user-growth indicators.
- **Senior Software Engineer (ML/LLM Systems) @ The New England Center for Children** March 2023 — March 2026
Built a transformer-based recommender that surfaces semantically related lessons across the curriculum (e.g., a "tomato" query returns not only direct lessons but adjacent skills like grocery shopping and burger preparation), with sub-second response. Led end-to-end deployment of a fine-tuned Llama 3.1 8B (LLaMA-Factory + LoRA, three rank iterations) for automated curriculum-resource generation, grounded via a RAG layer over NECC's internal PDF technical documentation; expert reviewers confirmed exact match with the curriculum's English voice.
- **Senior Data Consultant (Part-time) @ Sankio del Peru** September 2022 — June 2025
Multi-discipline consulting (data science, data engineering, analytics) for Peruvian clients across banking, logistics, and food services. Built classical ML models for client-specific use cases, orchestrated production ETL pipelines, and delivered BI dashboards driving stakeholder decisions.
- **Lead Data Engineer @ Teamcore** July 2022 — January 2023
Built AWS-based ETL pipelines (Glue, Athena, Lambda, S3) ingesting product-availability data from major retailer B2B platforms — Cencosud, Falabella, Oxxo, Soriana, Walmart. Led a team of five engineers and coordinated with Product on client requirements.
- **Backend Developer / Data Scientist @ Apurata** November 2021 — July 2022
Maintained core platform services for Apurata and aCuotaz — bank connections, transactional queues, back-office systems, and request-latency optimization. Transitioned to the data team to train and evaluate traditional ML models (XGBoost, Logistic Regression) for default-risk classification, scored on KS / Gini / AUC; built Metabase dashboards for stakeholder visibility.
- **Full-stack Engineer @ Barrick Gold Corporation** May 2021 — Nov 2021 — Engineered a full-stack risk management platform with data visualization dashboards; led network analysis for the post-acquisition company migration.
- **Research Intern @ Binghamton University** January 2021 — August 2021 — Analyzed a corpus of 80,000 COVID-19 texts with novel NLP approaches; findings published at CSoNet 2021.
- **HPC Semi-Senior Developer @ Everis** Feb 2020 — Dec 2020 — Engineered raw-material microservices for **INDITEX**; refactored the **Ministerio de Justicia (España)** certificate system to process 500k+ records.
- **Full-stack Developer @ DAVOS SALUD** July 2019 — December 2019 — Built a clinic-management web app (records, appointments, billing); deployed fastText sentiment analysis on social media data.

Skills

- **LLM & ML Systems:** PyTorch, TensorFlow, Hugging Face Transformers, vLLM, LLaMA-Factory, Unsloth, LangChain, LangGraph, LlamaIndex, MCP (Model Context Protocol / FastMCP), OpenAI API, scikit-learn, NumPy, Pandas, CUDA, faster-whisper
- **Data & Cloud:** Spark/PySpark, Airflow, Prefect, Databricks (Delta Lake, MLflow, Feature Store, Model Serving), AWS (SageMaker, Glue, Redshift, Athena, Lambda, S3, EC2, DynamoDB), GCP (BigQuery, VertexAI, Dataflow, Pub/Sub, Cloud Run, Looker)
- **Databases:** PostgreSQL (pgvector), MSSQL, MySQL, MongoDB, Redis, Valkey, Neo4j, OracleDB
- **Backend & Web:** FastAPI, Flask, Django, Node.js, TypeScript, React.js, Next.js
- **Languages & Tooling:** Python, SQL, TypeScript, Julia, C++, Java, Bash; Docker, Kubernetes, Git, GitHub Actions, Jenkins
- **Observability:** Prometheus, Grafana, Datadog, Sentry, Langfuse

Education

- **MSc. Computer Science**
University of London, September 2023 - Present
- **B.S. Software Engineering** — GPA 4.0
Universidad Privada del Norte, March 2016 - December 2020

Research and Projects

- **MichiBoost.jl: Pure Julia gradient boosting with native categorical feature handling and symmetric decision trees.**
Personal — Open-source project
Ongoing project — [Source](#)
- **DearDiary.jl: A machine learning experiment tracking tool for Julia**
JuliaAI - Open-source project
Ongoing project — [Source](#)
- **MLJ.jl: A Julia machine learning framework**
JuliaAI - Open-source project
Ongoing project — [Source](#)
- **Empowering Julia-based Data Science with MLflow**
Google Summer of Code — Open-source project
May 2023 — September 2023 — [Source](#)
- **Understanding the Issues Surrounding COVID-19 Vaccine Roll Out Via User Tweets**
The 10th International Conference on Computational Data and Social Networks — Conference paper
December 04, 2021 — doi: 10.1007/978-3-030-91434-9_18
- **Convolutional Neural Networks in Respiratory Conditions Detection: A Systematic Review of the Last 4 Years**
Summer School of Machine Learning at Skoltech — Research poster
August 16, 2020 — [Source](#)

Volunteering & Awards

- **REPU Seminar 2026 Judge**
Research Experience for Peruvian Undergraduates — April 2026
Served as a judge for the REPU Seminar 2026, evaluating undergraduate research presentations across various scientific disciplines and providing constructive feedback to participants.
- **Reviewer**
IEEE — July 2024
Reviewed academic papers for the 2024 IEEE Intercon conference, assessing submissions for technical accuracy, originality, and alignment with conference themes.
- **WHO COVID-19 Research Database Contributor**
World Health Organization — August 2022
Authored research paper "Understanding the Issues Surrounding COVID-19 Vaccine Roll Out via User Tweets," indexed in the WHO COVID-19 Research Database.
- **REPU Alum**
Research Experience for Peruvian Undergraduates — October 2020
Chosen as one of the REPU Computer Science 2021 undergraduate participants, working under the mentorship of Gisella Bejarano in Professor Arti Ramesh's research lab at Binghamton University.
- **Top 10% Spot, SMILES 2020**
Skolkovo Institute of Science and Technology — August 2020
Awarded an exclusive presentation spot (top 10% of 2,000+ applicants) to present the work "*Convolutional Neural Networks in Respiratory Conditions Detection: A Systematic Review of the Last 4 Years*".
- **OAS Cybersecurity Scholarship Recipient**
Organization of American States — May 2019
Awarded the opportunity to participate in the "*Creando una Trayectoria Profesional en Seguridad Digital*" program, receiving specialized training at Universidad Federico Villarreal in Lima.

Courses & Certifications

- **R Bootcamp**
Data Science Research Peru
March 2023 — [Certificate](#)
- **Scrum Foundation Professional Certificate SFPC™**
CertiProf
February 2023 — Certificate ID 83527332
- **Deep Learning with Flux.jl**
JuliaAcademy
February 2023 — Certificate No. cert_my5hq86p
- **Data Analytics for Ingestion and Validation**
Google Cloud Labs (sponsored by Teamcore)
December 2022
- **Summer School of Machine Learning**
Skolkovo Institute of Science and Technology
August 2020 — [Certificate](#)
- **Deep Learning Specialization**
deeplearning.ai
September 2020 — [Certificate](#)
- **Machine Learning**
Stanford University
November 2019 — [Certificate](#)
- **Creation of a Professional Trajectory in Digital Security**
Organization of American States (OAS)
May 2019 — [Certificate](#)