Santiago Martínez Novoa

Machine Learning Researcher | Multilingual NLP & Trustworthy LLM Evaluation

s.martinezn@uniandes.edu.co | LinkedIn | GitHub | Website

Education

Universidad de los Andes — Bogotá, Colombia
M.Sc. Information Engineering — *2025 – 2026 (To be conferred)*

Universidad de los Andes — Bogotá, Colombia
*B.Sc. Systems and Computing Engineering (**Cum Laude**), Minor in French* — *2021 – 2024*

Experience

Mitacs Globalink Research Intern - Collaborating Researcher — May 2026 – Present
Ontario Tech University — *Oshawa, Canada*

- Co-authored a Findings of EMNLP 2026 paper on adaptive language sampling; leading its follow-up on when to stop cross-lingual realignment, under review at ARR
- Co-developing M2-ALIGN: extending vision-language models to 11 languages while preserving their multilingual reasoning

Data Scientist — Mar 2026 – Present
EdTech Node, Universidad de los Andes — *Bogotá, Colombia*

- Lead the unit's adoption of data governance and quality policies, enabling university-wide data products such as academic percentiles and dropout indicators
- Lead the team's research line: dropout-prediction ML models behind 2 peer-reviewed publications (AIED & WEEF 2026)
- Manage the data behind "No Estás Solx," a peer-support platform serving 56,000+ students; the team received the 2026 Dejar Huella Award

Graduate Teaching Assistant — Jun 2025 – Oct 2026
Universidad de los Andes — *Bogotá, Colombia*

- Designed the hands-on curriculum for two sequential master's courses on RAG systems and AI agents, reaching 200+ students across 3 semesters
- Built tutorials, projects, and RAGAS-based evaluation pipelines with LangChain, LangGraph, and LangSmith

Data Scientist - Graduate Research Assistant — Jan 2025 – Mar 2026
Vice Dean's Office of Student Affairs, Universidad de los Andes — *Bogotá, Colombia*

- Replaced external counseling platform, saving the university ~$100K USD annually
- Co-implemented system alerting over 1000 critical student cases
- Implemented complete medallion architecture in Databricks, fully automating institutional data pipelines
- Optimized ETL pipeline performance by 87% by migrating from Pandas to Polars, reducing processing time from 15 to 2 minutes

Generative Artificial Intelligence for Research - Research Assistant — Jan 2025 – Jun 2025
Universidad de los Andes — *Bogotá, Colombia*

- Designed an LLM-as-judge evaluation architecture that grades student work against rubrics
- Built locally deployed RAG systems and custom AI agents for domain-specific research questions

Machine Learning Intern — Aug 2024 – Dec 2024
Precia PPV S.A — *Bogotá, Colombia*

- Developed MLP model reducing financial calculation time by 70% for client applications
- Presented our work on models and AWS infrastructure at Analytics Forum 2025
- Implemented end-to-end AWS infrastructure with automated ETL pipelines using Glue and Lambda

Summer Research Intern — Jun 2024 – Aug 2024
College of Computing and Information Science, Cornell University — *Ithaca, United States*

- Designed Data Discovery architecture to explore and modify datasets in various formats

- Developed upload and scoring system through semantic search across multiple databases

Technical Skills

- *Programming Languages*: Python, SQL, JavaScript, TypeScript
- *Machine Learning & AI*: Scikit-learn, Keras, PyTorch, TensorFlow
- *LLM & Natural Language Processing*: Transformers, BERT, LangChain, LangGraph, LangSmith, Ollama
- *Data Processing*: Pandas, Polars, Databricks
- *Cloud & DevOps*: AWS, Azure Functions, Azure Blob Storage
- *Databases*: PostgreSQL, MongoDB, ChromaDB, Pinecone