Santiago Martínez Novoa

Systems and Computing Engineer

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

RESEARCH INTERESTS

My research lies at the intersection of multilingual natural language processing, multimodal AI systems, and rigorous evaluation methodologies. I am particularly interested in developing techniques for effective cross-lingual knowledge transfer in low-resource settings, as well as in designing architectures that integrate textual and visual information across diverse linguistic contexts. I have worked on the comparative evaluation of highly-trained domain-specific models and Large Language Models for Spanish medical NLP tasks, exploring their behavior under zero-shot and few-shot learning settings. In computer vision, I specialize in image segmentation and classification for medical applications. As a systems and computer engineer, I am committed to building data-oriented tools and production-level Machine Learning systems to enhance student support in higher education, with an emphasis on real-world impact, interpretability, and data-driven decision-making.

EDUCATION

Universidad de los Andes   Bogotá, Colombia
M.Sc. Information Engineering   *2025 – 2026 (To be conferred)*
    GPA: 4.68/5.00

Universidad de los Andes   Bogotá, Colombia
B.Sc. Systems and Computing Engineering   *2021 – 2024*
    GPA: 4.58/5.00, ***Cum Laude***
    Minor in French

PUBLICATIONS

Martínez Novoa, S. "Detection and Segmentation of Malignant Melanoma Regions in Dermoscopic Images Using Machine Learning". *Undergraduate thesis presented at Universidad de los Andes to obtain the degree of Systems and Computing Engineer.* Universidad de Los Andes. Available at: Uniandes Repository (2024)

RESEARCH EXPERIENCE

FLAG Research Lab - A.I. and Machine Learning, Universidad de los Andes   2025-01 – Present
Evaluation and Enhancement of Natural Language Models in spanish   *Bogotá, Colombia*

- Research on improving natural language model capabilities for Spanish medical contexts, focusing on text classification and named entity extraction tasks under Professor Ruben Manrique Ph.D.
- Develop and train specialized models from scratch for specific Spanish medical domain tasks.
- Conduct comparative evaluations between highly-trained models and Large Language Models (LLMs) using prompting techniques such as zero-shot and few-shot learning.

Augmented Artificial Intelligence Initiative, Universidad de los Andes   2025-01 – 2025-06
Generative Artificial Intelligence for Research - Research Assistant   *Bogotá, Colombia*

- Focused on optimizing university research processes through custom AI agent development and automated evaluation systems.
- Built locally-deployed RAG systems using Ollama models and integrated Gemini API calls to answer domain-specific research questions.
- Designed LLM evaluation architecture to assess response quality and accuracy across different research contexts

using Prometheus, BERT score and OpenAI API.

- Automated student assessment processes by developing LLM-based evaluation systems that grade responses against established rubrics and expected answers.

College of Computing and Information Science, Cornell University 2024-06 – 2024-08
Summer Research Intern *Ithaca, United States*

- Designed a Data Discovery architecture to explore and modify datasets in various formats under Professor Sainyam Galhotra Ph.D.
- Developed the first version of the architecture including the upload and scoring through semantic search of multiple databases.

WORK EXPERIENCE

Universidad de los Andes 2025-06 – Present
LLM-Based Applications - Research Assistant *Bogotá, Colombia*

- Developing educational materials and practical exercises for a master's-level course focused on large language model applications.
- Created tutorials and course materials using frameworks including LangChain, LangGraph, and LangSmith for LLM and agent development.
- Designed practical exercises and assessments incorporating evaluation metrics such as RAGAS to measure LLM application performance.

Vice Dean's Office of Student Affairs, Universidad de los Andes 2025-01 – Present
Data Scientist - Graduate Research Assistant *Bogotá, Colombia*

- Process academic data and develop Power BI dashboards under the guidance of Professor Marcela Hernández Hoyos, Ph.D. and Professor Juan Pablo Reyes Gómez, Ph.D., delivering actionable insights on student performance and institutional metrics.
- Designed and developed Power BI dashboards to deliver actionable insights into student history, performance trends, and institutional matters.
- Optimized ETL pipeline performance by 87% (reduced processing time from 15 to 2 minutes) by migrating from Pandas to Polars for data retrieval and storage from Azure Blob Storage.
- Enabled the university to save approximately $100K USD annually by replacing an external academic counseling platform.
- Co-implemented a risk detection system that alerts stakeholders about over 1000 critical student cases.
- Designed and implemented the university's medallion data architecture, establishing a standardized and organized access point for institutional information.

Precia PPV S.A 2024-08 – 2024-12
Machine Learning Intern *Bogotá, Colombia*

- Co-develop and train a Multi-layer Perceptron (MLP) to reduce the calculation time of a critical financial metric by 70%, providing clients with significantly faster access to preliminary estimates.
- Design and implement end-to-end AWS infrastructure including robust ETL pipelines integrating multiple data sources (databases, FSx servers) and automate model deployment using AWS Glue and Lambda functions.
- Conduct data exploration and engineering to optimize model accuracy and efficiency across development and production environments.
- Presented our work at the Analytics Forum 2025 on model optimization and infrastructure architecture, including the use of MLPs, AWS pipelines, and data transformation strategies.

Vice Dean's Office of Student Affairs, Universidad de los Andes 2024-08 – 2024-12
Data Scientist - Undergraduate Research Assistant *Bogotá, Colombia*

- Co-implemented the ETL process (initially in Pandas, later migrated to Polars) to retrieve data from Azure Blob Storage, structure it, and store it in a relational database.
- Developed and conducted integrity tests to ensure the coherence and accuracy of a multidimensional model, validating against the original data.

Digital Transformation, Universidad de los Andes   2023-09 – 2024-05
Backend Developer   *Bogotá, Colombia*

- Applied expertise in Azure Functions and Python to actively contribute to software development projects, particularly focusing on seamless interaction with the Open AI API.
- Created comprehensive documentation for the project's infrastructure, emphasizing clarity and transparency. This initiative facilitated smoother collaboration among team members and enhanced the overall understanding of the software architecture.

Research Center of the Faculty of Engineering, Universidad de los Andes   2023-08 – 2023-12
Frontend Developer   *Bogotá, Colombia*

- Co-developed and implemented responsive and user-friendly frontend solutions for a software project at the university using Angular.
- Implemented dynamic access controls, tailoring user permissions based on roles and responsibilities.

Digital Transformation, Universidad de los Andes   2023-06 – 2023-08
Cross-Cultural Liaison and Software Developer   *Bogotá, Colombia*

- Facilitated effective communication between directors and 5 exchange students by providing translation services.
- Collaborated on software development projects utilizing Open AI and Azure Functions, successfully integrating the developed software into an interactive chatbot.

TEACHING EXPERIENCE

Systems and Computing Engineering Department, Universidad de los Andes   2024-01 – 2024-12
Undergraduate Research Assistant (CupiTaller Programming Support Center)   *Bogotá, Colombia*

- Supervised approximately 40 tutors and 70 teaching assistants as part of the 7-person team overseeing operations at CupiTaller, the university' programming support center, under Professor Eduardo Rosales, Ph.D.
- Designed programming projects for 'Introducción a la Programación', a course with +1100 students each semester.
- Cleaned, enriched, and processed datasets essential for each project, ensuring data accuracy completeness and consistency.

Systems and Computing Engineering Department, Universidad de los Andes   2021-08 – 2023-08
Undergraduate Teaching Assistant - Programming Tutor (Programming Support Center)   *Bogotá, Colombia*

- Provided 5 hours of weekly Python programming tutoring at the University's programming help center.
- Instructed more than 100 students, helping them develop fundamental coding skills in Python.

Universidad de los Andes   2022-01 – 2022-12
Undergraduate Teaching Assistant (Introduction to Programming)   *Bogotá, Colombia*

- Provided support to the professor during programming teaching laboratories, assisting approximately 40 students each semester.
- Corrected and graded basic programming projects.

Universidad de los Andes   2023-01 – 2023-06
Undergraduate Teaching Assistant (Fundamentals of Technological Infrastructure)   *Bogotá, Colombia*

- Corrected and graded tests, activities, and assessments
- Guided approximately 40 students to a thorough understanding of Fundamentals of Technological Infrastructure

HONORS AND AWARDS

Cum Laude   2025-04
Universidad de los Andes

- A Cum Laude degree is awarded to students with a GPA in the highest 3% GPA of all the students that have graduated from Universidad de Los Andes in a given faculty in the last five years.

Summer Undergraduate Research Fellowship (SURF)   2024-06
Cornell University

- Financial support awarded to participate in the Summer Undergraduate Research Fellowship (SURF) program at Cornell University

INVITED TALKS

Accelerating Credit Value Adjustment with Machine Learning. *Analytics Forum 2025 - Universidad de los Andes*, Bogotá, Colombia. 2025-05-09. [Web page]

SKILLS

Languages

- Spanish (Native)
- English (Advanced)
- French (Conversational)

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