

Rafael Mateus Carrion

rafaelmateuscarrion@gmail.com | rmateusc.github.io | linkedin.com/in/rmateusc/

EDUCATION

New York University - NYU (New York, NY, USA)

Sep. 2025 - May 2027

M.Sc in Data Science

- Awarded the **CDS Pathbreaker Scholarship** granted to only 2 students annually ($\sim 1\%$ of cohort).

Universidad EAFIT (Medellín, Colombia)

July 2018 - June 2023

Bachelor in Mathematical Engineering, Minor in Computer Science and Statistics

GPA:(4.38/5.0)

- **Proud of:** Offered a **Software Engineering Internship at Meta** (London, UK. Fall 2022).
- Selected as one of 40 participants across Latin America for Mercado Libre's **Machine Learning Bootcamp**.
- **Master's courses:** Machine Learning, Bayesian Statistics, Robust and Non-Parametric Statistics, Time Series.

CONFERENCE PAPERS

- D. Otero, **R. Mateus**, R. Balestriero, "Anomaly Detection In The Wild: Can SSL Handle Strong Distribution Imbalances?", in *Proceedings of the **NeurIPS 2024** Workshop of Self-Supervised Learning - Theory and Practice*, and in *Proceedings of the LatinX in AI Workshop at **NeurIPS 2024** (Awarded Oral Presentation and Best Poster Award)*. Vancouver, Canada, December 10-15, 2024. arxiv:2410.04289.
- D. Otero and **R. Mateus**, "Self-Supervised Learning for Identifying Defects in Sewer Footage", in *Proceedings of the LatinX in AI (LXAI) Research at **ICML 2024***. Vienna, Austria, July 21-27, 2024. arxiv:2409.02140
- M.S. Uribe, M. Franco, L.F. Londoño, P. Escudero, S. Álvarez, **R. Mateus**. "Comparing Three Agent-Based Models Implementations of Vector-Borne Disease Transmission Dynamics", in *Advances in Computing. CCC 2023. Communications in Computer and Information Science*, vol 1924. Springer. 2024. doi: 10.1007/978-3-031-47372-2.27.
- P. Escudero, M. Franco, M. S. Uribe, S. Álvarez and **R. Mateus**. "Using a Hybrid ABMS to Study the Propagation of Vector-Borne Diseases in an Urban Area with Heterogenous Geospatial Conditions", in *2023 Winter Simulation Conference (WSC)*. IEEE. pp. 3142-3153, 2023, doi: 10.1109/WSC60868.2023.10407951.
- M.A.C. Manrique, D. Otero Gomez, H. Laniado, **R. Mateus C**, et al., "Housing-Price Prediction in Colombia using Machine Learning", in *Proceedings of the 11th International Congress of Computing of Northern Chile (INFONOR)*, Atacama, Chile, Sept. 2-4, 2020.
- M.A.C. Manrique, M. Toro, D. Otero Gomez, **R. Mateus C**, et al., "Career Recommendation System for School Seniors based on Government Standardized-Tests Results", in *Proceedings of the 11th International Congress of Computing of Northern Chile (INFONOR)*, Atacama, Chile, Sept. 2-4, 2020.

WORK EXPERIENCE

ML Engineer at AI Fund

Mar. 2025 - Aug. 2025

LLM-Based Retrieval and Graph Reasoning Systems

- Part of the techinal team at Andrew Ng's Venture Studio, gaining exposure to cutting-edge applied AI research within an entrepreneurial ecosystem.
- Designed and implemented advanced retrieval pipelines integrating **vector databases** and **knowledge graphs** to improve document-centric question answering and reasoning.
- Developed and benchmarked state-of-the-art paradigms including **RAG**, **Graph-RAG**, **Think-on-Graph**, and task-specific LLM fine-tuning for single- and multi-hop QA.
- Engineered an end-to-end graph-based pipeline for **document ingestion**, **entity extraction**, and **knowledge graph construction** to enable semantic and graph-structured reasoning.

AI/ML Engineer at EXIT83 Consulting

Aug. 2023 - Jan. 2025

Self-Supervised Learning for Identifying Maintenance Defects in Sewer Footage

- Achieved strong performance on the Sewer-ML dataset with just 10% of the data for fine-tuning—significantly reducing annotation requirements. Additionally, trained a model that is one-fifth the size of current approaches, making it well-suited for small-device deployment and real-time detection in resource-constrained environments..
- **Paper accepted** at *LatinX in AI* Workshop at **ICML 2024** (Vienna, Austria).

- Participated in **NeurIPS 2024**, presenting a poster at the *Self-Supervised Learning - Theory and Practice* Workshop. Delivered an **Oral Presentation** and received the **Best Poster Award** at the *LatinX in AI* Workshop.

Private Equity Summarization Agent

- Developed a chatbot to interact with internal data sources of a private equity firm, enabling efficient information extraction and summarizing them into a one-page framework tailored to stakeholders' needs.
- Implemented a Retrieval-Augmented Generation (RAG) system using Pinecone databases and OpenAI embedding models, with a simple user-friendly interface built on Gradio for seamless end-user interaction.

Gesture Control on Low Resolution Edge Devices

- Developed a low-compute neural network tailored for gesture recognition on ToF sensor.
- Implemented the data collection and curation process. Developed a feedback loop that boosted model performance by integrating challenging samples into the training dataset.

HR Talent Matching Algorithm

- Developed a system to connect job opportunities with top talent by aligning optimal embeddings for both candidates and job profiles.
- Implemented a hybrid approach using dense and sparse vector representations, facilitating both semantic and lexicographical searches for more accurate and comprehensive talent-job matching.

AI/ML Specialist Intern at EXIT83 Consulting

Apr. 2023 - Jul. 2023

- Implemented anomaly detection solutions for complex systems by leveraging wavelet transforms and statistical models, improving system monitoring and predictive maintenance
- Developed and deployed an imbalanced text classification model by fine-tuning an OpenAI model on Azure.

Data Scientist and Analytics Intern at Nutresa S.A.

Jan. 2022 - Jul. 2022

- Collaborated with the Data Science team to develop and analyze machine learning models, gaining insights into customer behavior, market trends, and product characteristics.
- Managed databases and models using SQL and Python in the AWS cloud.

VOLUNTEERING

General Co-Chair at LatinX in AI Workshop @ ICML 2025

Mar. 2025 - Jul. 2025

- Organized and led the workshop by overseeing planning and logistics, running weekly stand-ups, and supervising co-chairs to ensure its success.

Presentation Co-Chair at LatinX in AI Workshop @ NeurIPS 2024

Aug. 2024 - Jan. 2025

- Managed the presentation day, scheduled and coordinated with keynote speakers, and oversaw all workshop activities.

Student Representative at EAFIT

May 2020 - Sep. 2021

- Mediated student assemblies, participated in internal university dialogues, and devised suitable solutions to address challenges arising from the 2021 Colombian National Strike.

Teacher Assistant, Mathematics Department at EAFIT

Jan. 2021 - Dec. 2021

Courses: Introduction to Python, Agent-Based Modeling and Simulation.

- Conducted weekly Python courses for beginners, covering basic libraries usage and data management. Mentored 40+ students in Agent-Based Modeling and Simulation, enhancing their conceptual and computational skills.

REVIEWING

- Reviewer at the Self-Supervised Learning - Theory and Practice Workshop at NeurIPS 2024.

SKILLS & INTERESTS

Languages: Python (Proficient), SQL (Good), Matlab (Good), R (Good), Java (Basic).

Software Tools: PyTorch, Wandb, Azure ML, Git, Sklearn, Numpy, Pandas, Matplotlib.

Languages: English (C1), Spanish (Native), French (Intermediate, B2), German (Basic, A2)

Hobbies: reading, running, boxing, cooking, and traveling.

Research Interests: Self-Supervised Learning, Cognitive Learning, Computer Vision.