Víctor Samuel Pérez Díaz

CONTACT INFORMATION

New York University
Courant Institute of Mathematical Sciences
60 5th Avenue, 5th Floor, New York, NY 10011

v.perez.diaz@nyu.edu
samuelperezdi.com

RESEARCH INTERESTS

I am interested in machine learning and computer vision in the sciences, with a focus on astronomy and physics.

EDUCATION

New York University, New York, NY
Courant Institute of Mathematical Sciences *January 2025 - Present*
MPhil, PhD Student in Computer Science
Advisor: David Fouhey

Universidad del Rosario, Bogotá, Colombia January 2018 - December 2021
Bachelor's in Applied Mathematics & Computer Science, *Focus on Artificial Intelligence and Data Science*
Thesis: *Unsupervised Machine Learning for the Classification of Astrophysical X-ray Sources (grade 5/5)*
Advisors: Rafael Martínez-Galarza & Alexander Caicedo-Dorado

PUBLICATIONS

5. **Pérez-Díaz, V. S.**, Kashyap, V., Ingram, J., Fouhey, D., Martínez-Galarza, J. R., Protopapas, P., Drake, J. J., Kim, D.W., and Garrafo, C., "The Chandra-Gaia Catalog of Counterparts: Resolving ambiguous Gaia matches to X-ray sources in the Chandra Source Catalog using Machine Learning", *In review. The Astrophysical Journal.*
4. Kitouni, O., Nolte, N., **Pérez-Díaz, V.S.**, Trifinopoulos, S. and Williams, M., "From Neurons to Neutrons: A Case Study in Interpretability", *Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024.*
3. **Pérez-Díaz, V.S.**, Martínez-Galarza, J. R., Caicedo Dorado, A. and D'Abrusco, R., "Unsupervised machine learning for the classification of astrophysical X-ray sources", *Monthly Notices of the Royal Astronomical Society, Volume 528, Issue 3, March 2024, Pages 4852–4871.*
2. **Pérez-Díaz, V. S.**, Trifinopoulos, S., "NuDyCLR: Nuclear Dynamic Co-Learned Representations", 2023, *2023 REYES Proceedings.*, 3.
1. **Pérez-Díaz, V. S.**, "Unsupervised Machine Learning for the Classification of Astrophysical X-ray Sources", 2021, *Universidad del Rosario Repository. Undergraduate Thesis Manuscript.*

RESEARCH IN PROGRESS

1. **Pérez-Díaz, V. S.**, Wang, R., Dissauer, K., Leka, K. D., and Fouhey, D., "Evaluating Solar Differential Emission Measure with Deep Learning", *In preparation.*

AWARDS & HONORS

Ser Pilo Paga Scholarship 2018 - 2021
Selected as a *Ser Pilo Paga* national scholarship recipient, which provides outstanding students from vulnerable populations in Colombia access to higher education in the nation's top universities.

Latin American Leadership Academy Scholarship 2019
Received a scholarship to attend the Mexico Leadership Bootcamp at Ciudad de México. I was the only colombian in the bootcamp, selected from more than 800 applicants.

Youth Ambassadors Program, U.S. Embassy in Colombia 2017
The Youth Ambassadors Program aims to promote mutual understanding, increase leadership skills, and prepare youth to make a difference in their communities. Part of a group of 10 young Colombians selected from more than 600 applicants.

Honor Mention in Classical Guitar, Conservatorio del Tolima 2017

Conservatorio del Tolima is one of the best educational institutions in music in Colombia. As a student of Classical Guitar in the music school, I received an honor mention requested by my guitar professor, Manuel Olaya.

RESEARCH EXPERIENCE

IT Specialist (Machine Learning) August 2024 - January 2025
Center for Astrophysics | Harvard & Smithsonian
Advisor: Vinay L. Kashyap

⇒ Developing a machine-learning based method for finding reliable counterparts to astrophysical sources detected by NASA's Chandra X-ray Observatory.

Lecturer and Researcher January 2024 - June 2024
Universidad del Rosario

⇒ Lecturer and Researcher at the School of Engineering, Science and Technology. Courses: Calculus.

Research Fellow July 2023 - January 2024
Harvard John A. Paulson School of Enginering and Applied Sciences (ongoing collab.)
Center for Astrophysics | Harvard & Smithsonian
Advisors: Pavlos Protopapas, Vinay L. Kashyap

⇒ Develop a machine learning based method to provide a better crossmatching between the Chandra Source Catalog 2.1 and the Gaia Data Release 3.

Research Mentee July 2023 - September 2023
REYES: Remote Experience for Young Engineers and Scientists
Massachusetts Institute of Technology (MIT)
Mentor: Sokratis Trifinopoulos, Michael Williams

⇒ Reviewed literature in multi-task learning and nuclear physics problems.
⇒ Implemented several dynamic loss weighting methods for a cutting-edge neural network model that predicts many nuclear properties with high accuracy.
⇒ Explored performance changes for each different weighting method. Derived conclusions and wrote a manuscript published in REYES Proceedings 2023.
⇒ Analyzed functional forms of the embedding space to derive corrections for nuclear physics models.

Research Collaborator January 2022 - July 2023
Universidad del Rosario
Advisor: Edgar Andrade-Lotero

⇒ Reviewed literature in modeling of group cognition using *El Farol* problem.
⇒ Implemented several models using Python, and simulated them numerous times to find the best hyperparameters.
⇒ Explored and expanded mathematical implications of models presented, particularly those based on Markov Fictitious Play.
⇒ Presented our results in the 3rd Meeting of Applied Mathematics and Statistics for Social Sciences and Economics.

Research Intern June 2022 - October 2022
Center for Astrophysics | Harvard & Smithsonian
Advisor: Rafael Martínez-Galarza
Collaborators: Alexander Caicedo-Dorado, Raffaele D'Abrusco

⇒ Performed several validation tests using confusion matrices and *bona-fide* regions, in order to analyze astrophysical relation between the classifications and locations.
⇒ Developed a final step for the pipeline in order to provide a final unique classification for all sources in our dataset. This was based on hard and soft voting classifiers using different detections of the same source.
⇒ Summarized the most important results and discussion in a manuscript, which will be my first paper also as a first-author. Paper submitted to Monthly Notices of the Royal Astronomical Society. Favorable review. Referee report in process.

⇒ Submitted and structured several first-author contributed talk posters.

Undergraduate Research Thesis January 2021 - November 2021
Universidad del Rosario
Advisors: Rafael Martínez-Galarza & Alexander Caicedo-Dorado

⇒ Reviewed literature in machine learning classification of astrophysical X-ray sources. Learned concepts in high energy astrophysics, particularly those related with the Chandra X-ray Observatory.
⇒ Structured a pipeline based on Gaussian Mixture Models and the Mahalanobis distance to classify sources of the Chandra Source Catalog, first using the power of unsupervised learning.
⇒ Implemented the proposed pipeline in 2000+ lines of code with Python, and tested it with the dataset numerous times in order to find the best hyperparameters and configuration.
⇒ Wrote a 57 pages-long dissertation introducing the field, the broad question, the specific research question and the most important findings. This work received the highest grade (5/5).
⇒ Presented a first-author Invited Talk at Chandra Data Science Workshop 2021.

FUNDING PROPOSALS

CfA Decadal Survey - White Paper 2022
Center for Astrophysical Machine Learning at CfA (now AstroAI). PI: Cecilia Garraffo, Co-Is: J. Rafael Martinez Galarza, Phillip Cargile, Joshua Wing, James F. Steiner, Vinay Kashyap, Aneta Siemiginowska, Lindy Blackburn, Peter K. G. Williams, Douglas Finkbeiner, Rosanne Di Stefano, Adam Foster, Avi Loeb, Morgan MacLeod, Floor Broekgaarden, Ethan Tregidga, **V. Samuel Perez Diaz**, and Edo Berger

Chandra X-ray Center Research Visitor Program - **$3k** 2022
Unveiling the nature of Chandra source: An unsupervised machine learning classifier for CSC 2.0 sources. Martínez-Galarza, R., **Pérez-Díaz, V. S.**

ADDITIONAL EXPERIENCE

Data Analyst February 2022 - August 2023
Propel

Propel identifies social organizations creating measurable impact and strengthen them through a Digital Transformation Fellowship. Our cohort-based approach builds a community for these organizations to thrive.

⇒ Conducting Research and Generating Insights from Data Collection.
⇒ Tracking Impact of Propel and the nonprofits in the portfolio.
⇒ Managing program assessments and generating feedback insights for improvement.
⇒ Monitoring databases and Salesforce CRM.

Frontend Developer and Data Co-Engineer July 2021 - March 2022
Universidad del Rosario

Transporte Limpio is a data platform that consumes real time measurements from IoT devices to estimate and visualize emissions of cargo vehicles in the city of Bogota. The Transporte Limpio project is managed by Bogotá Inteligente Alliance and funded by P4G. It corresponds to an institutional alliance between Universidad del Rosario, Universidad EAN, ANDI, Probogotá and Tigo.

⇒ Worked mainly as a Frontend Developer, collaborating in data engineering and analysis.
⇒ Developed algorithms for efficient loading of geographical data of emissions in the city.
⇒ Structured the initial versions of the design for the dashboard webpage, aimed towards the general public and private companies.
⇒ Wrote 5000+ lines of code using ReactJS for dashboard graphs and data processing for visualization.
⇒ Contributed in collaborative research and engineering discussions for the data architecture of the platform and the best models for estimating emissions.

Teaching Assistant August 2018 - December 2021
Universidad del Rosario

During my time in the program of Applied Mathematics and Computer Science, I assisted 100+ learning processes of undergraduate students in courses such as:

⇒ Logic, number and set theory *(4 periods)*

⇒ Differential and Computational Geometry *(1 period)*
⇒ Theory of Computation *(1 period)*
⇒ Calculus I and Calculus II *(3 periods)*
⇒ Mathematical Thinking *(1 period)*

SELECTED SCIENCE TALKS & POSTERS

Science Talks

NASA Hyperwall Talk at AAS 245 (National Harbor, MD) — January 2025
MIT IAIFI Summer School Lighting Talk (Cambridge, MA) — August 2024
CfA AstroAI Invited Talk (Cambridge, MA) — Dec 2023
3rd Meeting of Applied Mathematics and Statistics for Social Sciences and Economics (Remote) — May 2023
Dr. Pepi Fabbiano's CXC-DS Group Talk *(Remote)* — March 2022
Thesis Dissertation *(Universidad del Rosario)* — Nov 2021
Chandra Data Science Workshop Invited Talk *(Remote)* — Aug 2021
Dr. Pepi Fabbiano's CXC-DS Group Talk *(Remote)* — March 2021

Posters

Physics and AI Conference 2026 (Stanford, CA) — June 2026
CVPR 2026 (Denver, CO) — June 2026
AAS 245 (National Harbor, MD) — January 2025
MIT IAIFI Summer Workshop (Cambridge, MA) — Aug 2024
International Conference on Machine Learning - ICML2024 (Vienna, Austria) — July 2024
20th Divisional Meeting of the High Energy Astrophysics Division of the AAS *(Waikōloa, HI)* — March 2023
American Astronomical Society Meeting #241 *(Seattle, WA)* — Jan 2023
19th Divisional Meeting of the High Energy Astrophysics Division of the AAS *(Pittsburgh, PA)* — March 2022

SELECTED OUTREACH

Colegio Diversificado Inca English Day
Special Invited Speaker *(Spanish) (Barranquilla, Colombia)* — Nov 2023
Engaging dialogue and Q&A sessions with hundreds of curious middle-school students in the Colombian Caribbean, fostering scientific curiosity and knowledge sharing.

Reading the sky through the lens of engineering, science and technology
Special Invited Seminar *(Spanish) (Bogotá, Colombia)* — May 2023
Talk title: "Eyes in the sky: using machine learning in astrophysics" (50 min).
Delivered an inspirational talk to approximately 50 undergraduates at Universidad del Rosario, sharing my personal and research journey. I provided practical tips and motivation for pursuing a career in STEM, emphasizing the importance of perseverance and resilience in overcoming challenges to achieve their dreams.

ADVISING & MENTORSHIP

Gabriela Valencia Zuñiga, Universidad del Valle *(Colombia)* — 2024
(currently a PhD Student in Astronomy at Drexel University)
Mentored independently.

Ana María Garzón, Universidad del Rosario *(Colombia)* — 2020 - 2022
(currently in senior year of Applied Mathematics and Computer Science, ELAP Canada Alumni)
Mentored through the Mentoring Plan of the Student Council and then independently.

Nicolás Triana, SENA *(Colombia)* — 2019
(now Intern in Falabella Seguros, enrolled in a Technologist of Information Systems program.)
Mentored through *Más Allá de 11* Mentorship Program.

LEADERSHIP & SERVICE

College Leadership & Service

URosario ACM Student Chapter – Chair/Co-Founder — 2021 - 2022
Co-founded the Association for Computing Machinery Student Chapter at Universidad del Rosario, the first of its kind in the School of Engineering, Science and Technology. Led the call for joining the chapter, and the first URosario Programming Contest, in which ~ 30 students participated.

Student Council of Natural Sciences and Mathematics

President/Chancellor/Student Representative 2018 - 2020

Elected democratically as president (2020) by the students of the Faculty of Natural Sciences and the School of Engineering, Science and Technology at Universidad del Rosario. Represented ∼300 science students. Previously acted as Chancellor (2019) and Student Representative (2018). Led many interdisciplinary and student-focused initiatives, such as a Mentoring Plan for first year students, community celebrations of Pi Day with invited talks, extracurricular meetings with professors of the department and students, mental health and safety protocols for students, etc. 20-hr/wk.

Other Service

Latin American Leadership Academy (LALA) – Data Team Member Aug 2020 - Sep 2021

The LALA Data Team is in charge of collecting, cleaning, analyzing data and reporting insights gained. Contributed in discussions focused on better collection of data. Analyzed and presented insights on data from the alumni.

Más Allá de 11 – Mentor Aug 2019 - Jul 2020

Mentoring program seeking that students from schools in vulnerable areas of Bogotá can move away from their daily context in order to broaden their horizons and build a better future. I mentored student Nicolás Triana, when he was a 10th level high school student in Ciudad Bolivar neighborhood.

Latin American Leadership Academy (LALA) – Data Coach Jul 2020

Planning and execution of data collection strategies for Virtual Bootcamps, in order to have suitable data for analysis and report.

Partners of the Americas – Youth Ambassador Jul 2017 - Nov 2017

Intensive leadership and social sensibility training, with more than 72 hours of practice in NGOs from Washington D.C. and Tennessee.

SELECTED MEDIA FEATURES

Article – MIT Technology Review April 2024

Astronomers are enlisting AI to prepare for a data downpour. AstroAI article which references my paper in classification of X-ray sources.

Podcast Interview (Spanish) – InformalMENTES Jun 2023

Las MACC en la astrofísica hosted by the Office of Research and Innovation at Universidad del Rosario.

Podcast Interview (Spanish) – Enrólate con la U Aug 2022

From Harvard-Smithsonian, let's talk about science, technology, and AI hosted by journalism students at Universidad del Rosario.

Journal Article – Revista Divulgación Científica, Universidad del Rosario Dec 2020

I just want to be Samuel... to inspire and have an impact by **VS Pérez Díaz**, X Serrano Gil.

Article – NBC News Jan 2018

Amid U.S.-Venezuela tensions, a youth program builds ties, fosters understanding by Patricia Guadalupe.

Article (Spanish) – El Olfato Newspaper Oct 2017

Student of a school from Ibagué became a 'Youth Ambassador' in the United States

Article – NBC News Aug 2017

Go South! Youth Ambassador Program Fosters Ties, Connections to Latin America by Patricia Guadalupe.

Article – Partners of the Americas May 2017

2017 Colombian and Venezuelan Youth Ambassadors Selected

PROFESSIONAL SOCIETIES & GROUPS

International CHASC Astro-Statistics Collaboration 2023 - Present

AstroAI 2022 - Present

LatinX in AI (LXAI) 2022 - Present

NSF Institute for Artificial Intelligence and Fundamental Interactions (IAIFI) 2023 - 2025

ACM Student Member and Student Chapter Member 2021 - 2022

WORKSHOPS & COURSES

IAIFI Summer Workshop Aug 2024

(MIT)

Poster: Crossmatching Astronomical Catalogs with Machine Learning

IAIFI Summer School Aug 2024

(MIT)

Lighting talk: Crossmatching Astronomical Catalogs with Machine Learning

Learning Machine Learning Summer School — July 2021
(Universidad del Rosario)
Assisted the Summer School invited by Dr. Alexander Caicedo-Dorado with a Teaching Assistant scholarship.

Complex Networks — July 2020
(School of Applied and Industrial Mathematics CoSIAM 2020)
Assisted as an invited student.

VII National Convention of Leaders, Grantees, and Ex-Grantees — Aug 2018
(Centro Colombo Americano)
Assisted with an Alumni scholarship.

UN Mathematics Meeting V — June 2018
(Universidad Nacional de Colombia)

Big Data CO — June 2018
(Universidad del Rosario)
Assisted with an scholarship as an invited undergraduate student.

OTHER SKILLS & TRAINING

Software Tools & Packages
- TOPCAT, ESASKY, Aladin Sky Atlas, Git, RStudio, AWS, LaTeX, ReactJS, Visual Studio Code, Atom.

Programming Languages
- Python, Javascript, Java, C++, MATLAB, R, Processing.

OS Proficiency
- Linux *(Ubuntu, WSL, Fedora)*, Windows.

Languages
- Spanish (Native Proficiency), English (Native Proficiency), Portuguese (Limited Working Proficiency).

References available upon request.
Last updated: May 30, 2026.