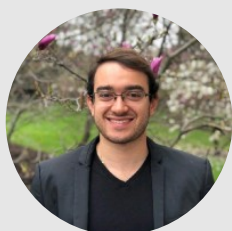


Andres F. Cano Botero

Researcher | Data Scientist



About me

I am a Data Scientist with a research background in quantum chemistry and plasma physics. My passion lies in exploring the synergy between data-driven methodologies and the natural sciences. Having witnessed the impact of AI and machine learning in computational chemistry, I am convinced that novel, physics-informed algorithms hold immense potential for simulating and extracting insights about our world. I am eager to contribute to this endeavor.

Contact

👤 From Manizales, Colombia

✉ andrescano30@gmail.com

☎ +1 404-956-0320

📍 Atlanta, GA

🌐 [andresfelipecano](#)

🔗 [andrescano](#)

Languages

🇪🇸 Spanish - Native Language

🇬🇧 English - Fluent

🇫🇷 French - Conversational Knowledge

🎓 EDUCATION

08/2020-

08/2022



M.S. Mechanical Engineering University of Colorado, Boulder

📍 Boulder, CO

Thesis: *Alkyl End Group Effects on the Thermal Decomposition of Fuel Additives.*

Advisor: Dr. Nicole J. Labbe

My work focused on optimizing and leveraging electronic structure theory methods and kinetic modeling to identify the decomposition products of two potential diesel additives. Studying the unimolecular decomposition of these molecules provides valuable insights into the formation mechanisms of soot particles in the exhaust of diesel engines.

GPA: 3.95/4.0

08/2015-

05/2019



B.S. Astrophysics University of Rochester

📍 Rochester, NY

Sigma Pi Sigma (National Physics Honor Society)

Lorenzo de Zavala Scholar, National Hispanic Institute

American Physics Society Member

💼 WORK EXPERIENCE

2022-

present

Data Scientist

📍 Atlanta, GA (Remote)

Dataiku

- Scoped, built, and deployed sophisticated production-level AI projects for customers across multiple domains, including natural language processing (NLP), optimization, computer vision, large language model (LLM) applications, and time series forecasting.
- Collaborated with R&D and product teams to develop custom Python-based plugins that enhance Dataiku's functionality.
- Led technical teams at non-profit organizations, such as the ALMA Observatory, that aim to develop data-driven solutions.
- Provided strategic input, coaching, and training to Fortune 500 companies across multiple industries (e.g., technology, pharmaceuticals, banking) seeking to implement AI solutions to optimize business processes.

2020-

2022

Graduate Researcher

📍 Boulder, CO

Labbe Lab, University of Colorado Boulder

- Leveraged university supercomputing resources and legacy software (written in C++ and Python) to build reaction mechanisms that uncover the thermal decomposition of soot-reducing diesel additives.
- Optimized research methods by writing a backward-compatible, high-performance computing pipeline that used Python and Unix shell scripts to generate, submit, and run quantum chemistry calculations on the RMACC Summit supercomputing cluster, saving days to weeks of computational work.

2019-

2020

Data Scientist

📍 Bogota, Colombia

Opera Urban Investment Firm OIU

- Developed a Streamlit app that displayed market trends, sales forecasts, and client activity from OIU's website. This established direct communication between internal teams and improved client outreach.

2018-

2019

Undergraduate Researcher

📍 Rochester, NY

Extreme State Physics Laboratory (XSPL), University of Rochester

- Built a high-voltage transformer to supply ± 200 kV to the High Amperage Driver for Extreme States (HADES), a 250 GW pulsed-power driver used to model matter at high temperatures and pressures.

📖 PUBLICATIONS







Journal Article **End Group Effects on the Thermal Decomposition of Oxymethylene Ethers**, AFC Botero, P. Shah, O Aguirre, J Sampathkumar, NJ Labbe, *Aimed for: Fuel* *in progress*

Journal Article **Temperature- and Pressure-Dependent Decomposition of Diethoxymethane and Dipropoxymethane Fuel Radicals.**, AFC Botero, P. Shah, O Aguirre, NJ Labbe, *Aimed for: PCCP* *in progress*

Page Summary

This page highlights the various projects I have worked on in both academic research and industry settings. It showcases my experience in developing automation tools for computational chemistry, building machine learning models for data science operations, and creating applications that streamline processes such as personal literature reviews and intricate MLOps pipelines. Each project demonstrates my ability to solve complex problems by leveraging computational tools, and together they paint a picture of my professional journey thus far.

Contact

-  From Manizales, Colombia
-  andrescano30@gmail.com
-  +1 404-956-0320
-  Atlanta, GA
-  [andresfelipecano](#)
-  [andresfcano](#)

RELEVANT RESEARCH & PROJECTS

Academic Projects - Labbe Lab

CU Boulder

Optimizing Computational Chemistry Methods
Motivation Accurately measuring the ground state energy is crucial for modeling biofuels at the molecular level. This requires exploring the molecule's potential energy surface through a hindered rotor analysis as part of the geometry optimization process to find the lowest energy conformer. Traditionally, this process demands significant manual work from computational scientists, who must visually identify rotors and prepare numerous files for supercomputing jobs.
Solution I developed a Python script suite that automates the identification of rotors in a given molecule and generates the necessary files for supercomputing cluster jobs. It also automatically creates input files for ab initio and Density Functional Theory (DFT) calculations, laying the groundwork for fully automating the computational process of calculating ground state energies. This automation significantly reduces manual effort, saving weeks of computational work in the long term.

CU Boulder

Unimolecular Decomposition of Diesel Additives
Motivation The higher compression ratios in diesel engines make them more prone to incomplete combustion, leading to the formation of harmful pollutants such as NOx. Efforts to reduce soot emissions have led to the exploration of fuel additives that blend well with diesel and minimize soot generation. Understanding the molecular dynamics and high-temperature and pressure decomposition of these additives is crucial for improving combustion efficiency and reducing pollutant formation.
Solution I conducted a study on the unimolecular decomposition of two potential diesel additives to gain insights into their behavior during combustion. I then conducted a deeper analysis of the decomposition of the primary radicals generated in the initial stages of combustion. The results revealed the key reactions that occur during the onset of combustion, providing valuable knowledge for developing cleaner diesel fuels. More details of these two studies can be found in my [thesis](#).

Industry Projects - Dataiku

Dataiku

Automating RFPs with Self-Hosted LLMs
Motivation: Field experts at a major insurance company manually handled hundreds of requests for proposals (RFPs), many of which had recurring themes, leading to inefficiencies and a loss in market competitiveness.
Solution: I designed and led the development of one of the first retrieval-augmented generation (RAG) pipelines for a customer at Dataiku that utilized open-source embeddings and large language models (LLMs) on local GPUs. By collaborating closely with stakeholders to ensure the solution met their needs, I built a Dash web application that enabled real-time querying, significantly improving RFP response times and overall efficiency.

Dataiku

AI for Science - ML for Observation Quality
Motivation After each science observation, astronomers on duty (AoD) at a large observatory need to assess the quality and usability of the data. This was done by manually reviewing data and categorizing it into three groups: Pass (usable data), Semi-Pass (data may be useful but needs to be repeated), or Fail (unusable data requiring a repeat observation). This process is time-consuming, especially for uncertain cases.
Solution To streamline this process, I spearheaded the development of a classification model to automate the assessment of observational quality. The model was robust and aligned with operational needs, allowing AoDs to focus only on uncertain cases. This solution greatly increased the efficiency of the data review process, reducing the manual workload.

Personal

Streamlining Literature Review with AI-Driven Analysis
Motivation: Conducting literature reviews is often a time-consuming process, especially when searching for specific methods or insights across multiple papers.
Solution: I developed a Python-based retrieval-augmented generation (RAG) application using Ollama as the backend and Streamlit for the front end. The application allows users to directly query large language models, significantly speeding up the literature search process and providing detailed responses with source citations.

– Soft Skills and Strengths –

- Curiosity
- Self Motivated
- Driven
- Independent Thinking
- Problem Solving
- Team Work
- Good Communication

Professional Skills

- Computational Physics & Chemistry
- Electronic Structure Modeling
- Data Science
- Machine Learning
- LLM Application Development
- MLOps
- Astrophysics & Mathematics
- Statistical Mechanics
- Engineering Design

Other Interests

- Guitar 🎸

• Chess ♟️

• Gym 🏋️

• Travel 🌍
- Triathlon 🏊

• Cooking 🍳

• Calisthenics 🤸

• Books 📖

💻 INFORMATION TECHNOLOGY SKILLS

Data Science and App Development	Python: <i>Higly Specialized</i> HTML/CSS/JavaScript: <i>Intermediate</i>
Version Control	Git: <i>Intermediate</i>
AI/ML	Sci-kit Learn : <i>Advanced</i> Hugging Face & Langchain : <i>Intermediate</i> PyTorch & Keras : <i>Intermediate</i>
Big Data	Spark : <i>Intermediate</i> SQL : <i>Intermediate</i> Kubernetes & Cloud Deployments : <i>Beginner</i>
Typesetting	L^AT_EX: <i>Intermediate</i>
Electronic Structure Theory Software	Gaussian16 : <i>Advanced</i>