# Andres F. **Cano Botero**
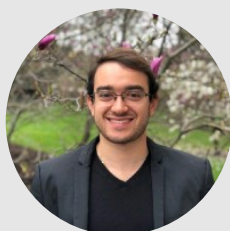## Researcher | Data Scientist

**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
- andresfcano

**Languages**

- Spanish - Native Language
- English - Professional Knowledge
- French - Conversational Knowledge

## 🎓 EDUCATION

**08/2020-08/2022**

**M.S. Mechanical Engineering**  ⚲ Boulder, CO
**University of Colorado, Boulder**
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 methodsopera 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**  ⚲ Rochester, NY
**University of Rochester**
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 in their work.

**2020-2022**

**Graduate Researcher**  ⚲ Boulder, CO
Labbe Lab, University of Colorado Boulder
- Leveraged CU 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**
*in progress*
**End Group Effects on the Thermal Decomposition of Oxymethylene Ethers**, AFC Botero, P. Shah, O Aguirre, J Sampathkumar, NJ Labbe, *Aimed for: Fuel*

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

─────── **About me** ───────

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

in andresfelipecano

○ andresfcano

## 🗂 RELEVANT RESEARCH & PROJECTS

### Academic Projects - Labbe Lab

#### Optimizing Computational Chemistry Methods   📍 CU Boulder

**Motivation:** Modeling biofuels at the molecular level requires exploring molecular rotors during geometry optimization to obtain an accurate calculation of the ground state energies of the lowest energy conformer. Traditionally, this involves slow, manual work by computational scientists.

**Solution:** I ideated and developed a Python script to automate rotor identification and file preparation for supercomputing jobs. The script was later expanded to generate input files for ab initio and density functional theory (DFT) calculations, streamlining the workflow and reducing manual effort. This automation saves weeks of computational work over time, improving efficiency in molecular modeling.

#### Unimolecular Decomposition of Diesel Additives   📍 CU Boulder

**Motivation:** Diesel engines are notorious for emitting harmful pollutants such as NOx. In response, research has focused on fuel additives that can effectively blend with diesel to reduce gas and soot emissions, thereby minimizing the pollutant output from these engines.

**Solution:** I led a study on the unimolecular decomposition of two potential diesel additives to gain insights into their behavior during combustion. This study was followed by 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 engine simulations of these fuels. More details can be found in my thesis.

### Industry Projects - Dataiku

#### Automating RFPs with Self-Hosted LLMs   📍 Dataiku

**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.

#### MLOps and CI/CD of Data Science Projects   📍 Dataiku

**Motivation:** A U.S. airline's data science team was developing multiple projects using Dataiku but lacked a structured machine learning operations (MLOps) process, which caused projects in development and production to be in the same environment. This led to security risks and inefficiencies.

**Solution:** I led the effort to design and implement a clear MLOps strategy that adhered to the company's CI/CD process. I spearheaded the development of an Azure DevOps pipeline in conjunction with the airline's Cloud Architecture and Security departments, automating project deployments and ensuring that security and data quality checks were in place. This solution not only streamlined project delivery but also enhanced transparency and operational efficiency across the organization.

#### AI for Science - ML for Observation Quality   📍 Dataiku

**Motivation:** After each science observation, astronomers on duty (AoD) at a major observatory had to manually assess the quality of the data, categorizing it as pass, semi-pass, or fail. This process was time-consuming, especially for uncertain cases.

**Solution:** I spearheaded the development of a machine learning model that automated the assessment of observational data 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.

#### Streamlining Literature Review with AI-Driven Analysis   📍 Personal

**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 Confidence

Self Driven   Problem Solving

Team Work   Constantly Learning

Good Communication

## —— Professional Skills ——

Computational Physics & Chemistry

Electronic Structure Modeling

Data Science   Machine Learning

LLM Application Development

ML Operations

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**: *Higly Specialized* <br> **HTML/CSS/JavaScript**: *Intermediate* |
| **Version Control** | **Git**: *Intermediate* |
| **AI/ML** | **Sci-kit Learn** : *Advanced* <br> **Hugging Face & Langchain** : *Intermediate* <br> **PyTorch & Keras** : *Intermediate* |
| **Big Data** | **Spark** : *Intermediate* <br> **SQL** : *Intermediate* <br> **Kubernetes & Cloud Deployments** : *Beginner* |
| **Typesetting** | **LaTeX**: *Intermediate* |
| **Chemical Kinetic Modeling and Simulation** | **Gaussian16** : *Advanced* |