

## Contact

+1 (608) 395 8543 (Mobile)  
cmcouto.silva@gmail.com

www.linkedin.com/in/cmcouto-silva  
(LinkedIn)  
cmcouto-silva.github.io (Personal)

## Top Skills

Transformers  
Chatbots  
Generative AI

## Languages

English (Professional Working)  
Portuguese (Native or Bilingual)  
Spanish (Limited Working)

## Certifications

Dataiku ML Practitioner  
Manipulation of NGS Data for  
Genomic and Population Genetics  
Analyses - 2nd edition  
The R Programming Environment  
Intermediate Machine Learning  
Machine learning in Python with  
scikit-learn

## Honors-Awards

Academic Honor for the best class  
performance  
Professor Bertha Lange de Morretes'  
Award — Second place in Poster  
Presentation at the 21st Congress of  
Biologists (CRBio-01)

## Publications

Selection scan reveals three new loci  
related to high altitude adaptation in  
Native Andeans  
Indigenous people from Amazon  
show genetic signatures of  
pathogen-driven selection

# Cainã Max Couto da Silva

Data Scientist | Machine Learning Engineer | AI Engineer | Gen AI |  
LLM | RAG | Agents | MLOps | Researcher  
São Paulo, Brazil

## Summary

I'm a data scientist and AI engineer with 10+ years of experience building data-driven solutions for academia, startups, and Fortune 500 companies. I have developed end-to-end AI applications, from research published in top scientific journals to production-ready ML systems grounded on real business impact.

I also speak at conferences and contribute to open-source Python projects.

## Key Achievements

- Postdoctoral Data Science Experience at UW-Madison (USA)
- MSc, PhD, and MBA at USP - ranked the best university in Latin America
- Worked for 2 Fortune 500 companies
- Best academic performance

## # Skills

- Programming languages: Python, SQL, and bash scripting
- Main frameworks: LangChain, LangGraph, LangSmith, Hugging Face, Pytorch, scikit-learn.
- Machine Learning: predictive models, time series, anomaly detection, clustering, etc.
- Deep Learning: model fine-tuning, computer vision for object detection, document intelligence.
- Generative AI: LLMs, AI agents, chatbots, RAG systems, prompt engineering.
- MLOps: model deployment with Docker, CI/CD pipelines, version control, APIs
- Big Data: Spark (PySparkSQL), BigQuery, Postgres, and MySQL.
- Cloud technologies: GCP (including Vertex AI) and AWS.
- Data Platforms: Databricks and Dataiku.

Mitigating Nonproductive Time:  
A Novel Algorithm for Dsl Fault  
Detection

Population Histories and Genomic  
Diversity of South American Natives.

I can successfully and efficiently translate business problems into technical solutions.

Open to collaborations, AI engineering roles, and thought partnerships.

Let's connect!

---

## Experience

University of Wisconsin-Madison

Postdoctoral AI Engineer

August 2024 - Present (11 months)

Madison, Wisconsin, United States

Deliveries:

- Building an agentic RAG-based text-to-SQL AI system that enables stakeholders and domain scientists to query live databases using natural language via Slack, boosting team productivity and contributing to AI Adoption.
- Developed SQLDeps: an open-source Python package leveraging LLMs to automatically extract table and column dependencies and outputs from complex SQL scripts 150X faster and >400X cheaper than human expert labor.
- Developed and optimized cattle mapping using cutting-edge deep learning models on high-resolution satellite imagery, providing actionable intelligence to combat Amazon deforestation.
- Engineered a pioneering machine learning classifier to assess data quality from farm properties at scale, implementing feature engineering on geospatial and entity data that enabled scalable data integrity assessment for the first time.
- Lead UW-Madison students in research projects, mentoring on machine learning and computer vision techniques.
- Automated database workflows, reducing manual intervention and boosting team productivity.
- Delivered private data analysis reports for stakeholders in Brazil and the USA.

Tech Stack: Python, PyTorch, LangChain, LangGraph, LangSmith, PostgreSQL, PGVector, FastAPI, Pydantic, Slack SDK, Docker, Git, UV, Streamlit, LLM providers.

## SLB

Data Scientist / Machine Learning Engineer

January 2023 - July 2024 (1 year 7 months)

Houston, Texas, United States

I developed end-to-end AI SaaS products for our internal customers.

### Main activities:

- Building predictive models using either statistical or machine learning approaches to assess the health of the company assets (tools)
- Assessing the technical feasibility of new projects through data analysis

### Deliveries:

- Deployed multiple machine learning models to predict asset failures and optimize maintenance schedules:
  - Designed explainable risk indices using regression and clustering
  - Published a statistically based model for detecting failures (see paper)
- A custom-trained OCR model to extract dimensions of engineering drawings
- Trained models on Azure Document Intelligence to extract target components from vendor documents.
- Led an idea for innovation using AI that ranked in the top 12 out of 237 teams worldwide and pitched it to C-levels.

### Quick facts:

- Together, savings from deliveries sum up to hundreds of thousands of dollars.
- I worked in one of the most diverse teams in our company, interacting with people from the US, India, Europe, and South America in my daily routine.

Tools: Dataiku, GCP, SQL, Python, PyTorch, Streamlit, machine learning libraries (e.g., scikit-learn), data visualization libraries.

## dnc.group

Data Science Consultant (Education)

October 2021 - July 2024 (2 years 10 months)

São José dos Campos, São Paulo, Brazil

I have worked in multiple roles: facilitator, mentor, and consultant/instructor.

As a consultant/instructor, I prepared the course materials and recorded classes. I recorded four modules: statistics, data cleaning/wrangling, clustering, and model deployment. I have also devised data science activities about descriptive and inferential statistics, unsupervised machine learning models, MLFlow, big data with PySpark, among others.

As a mentor, I assessed student reports and addressed their questions through Q&A sessions, helping them with academic and real-world projects.

As a facilitator, I participated in data science activities from exploratory analysis to model deployment. I also prepared some of those activities.

## Ambev Tech

### Data Scientist

May 2022 - December 2022 (8 months)

São Paulo, São Paulo, Brazil

Ambev Tech is the tech hub of AB InBev, the largest brewery company in the world.

### Squad: Revenue Management

#### Activities / Deliverables:

- Rule-based automation of pipelines for price engines using PySpark
- Forecasting prediction:
  - Statistical model to identify price change points for several SKU categories
  - Monthly time series modeling of selling volume using state-of-the-art forecasting methods like hierarchical reconciliation
- Training for data scientists

Tools: Python, PySpark, MLFlow, Scikit-learn, Pycaret, Statsmodels, Forecasting frameworks (Prophet, NeuralProphet, StatsForecast)

## Remessa Online

### Data Scientist

October 2021 - April 2022 (7 months)

São Paulo, São Paulo, Brazil

#### Activities:

- Advanced statistical modeling: data exploratory analysis, hypothesis testing, time series forecasting, predictive analyses (e.g., regression and classification), customer and product segmentation, etc.

Deliverables:

- Multiple time series forecasting for customer segments (weekly and monthly)
- Model for predicting the probability of the customer recurrence
- In-depth study of churn analysis through inferential statistics

Tools:

- Python • PySpark • MLFlow • Scikit-learn • Pycaret • Catboost • Prophet • AWS • GitHub • Data visualization libraries (Plotly, Seaborn, Matplotlib)

## Eli Lilly and Company

### Safety Data Sciences Associate

June 2021 - October 2021 (5 months)

Indianapolis, Indiana, United States

Activities:

- Queries and reports for worldwide company members

## Universidade de São Paulo

### Doctoral Researcher

July 2016 - April 2021 (4 years 10 months)

São Paulo, São Paulo, Brazil

Activities:

- Data analysis, visualization, and scientific reporting of genetic data using R, Python, and bash scripting.
- Application of non-supervised algorithms (e.g., PCA), descriptive and inferential statistics.

Deliverables:

- Internal R package with customized functions to facilitate multiple analyses
- Three scientific papers published in top-tier international journals

I presented our preliminary work at international conferences, including in the USA. I also took a three-month internship in Barcelona, Spain.

## Universidade de São Paulo

### MSc Researcher

April 2014 - March 2016 (2 years)

São Paulo, São Paulo, Brazil

I studied the role of the interaction between the proteins PrPC and STIP1 in adult neurogenesis.

Main techniques: primary cell culture, immunofluorescence, and hypothesis testing.

Universidade de São Paulo  
Undergraduate Research Fellow  
January 2013 - December 2013 (1 year)  
São Paulo, São Paulo, Brazil

UNG- Universidade Guarulhos  
Undergraduate Fellow  
July 2011 - December 2012 (1 year 6 months)  
Guarulhos, São Paulo, Brazil

---

## Education

MBA USP/Esalq  
Data Science & Analytics, Statistics · (May 2021 - August 2023)

Universidade de São Paulo  
Doctor of Philosophy - PhD, Genetics and Evolutionary Biology · (July 2016 - April 2021)

Universidade de São Paulo  
Master of Science - MS, Biological Sciences · (April 2014 - March 2016)

Universidade Guarulhos  
Bachelor's degree, Biology/Biological Sciences, General · (February 2011 - December 2013)