Cainã Max Couto da Silva

Senior AI Engineer | GenAI | AI Agents | RAG | MLOps

São Paulo, Brazil

+1 (608) 395 8543 | ✉ cmcouto.silva@gmail.com | 🔗cmcouto-silva | ⓖcmcouto-silva

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've developed end-to-end AI applications, from research published in top scientific journals to production-ready ML systems with measurable business impact. I also speak at conferences and contribute to Python open-source projects.

My current focus is building LLM-powered agentic AI applications that boost productivity and drive real-world value. I was selected to present a 4-hour tutorial at SciPy 2025 on building an AI agent for natural language to SQL query execution on live databases.

EXPERIENCE

Postdoctoral AI Engineer
Aug 2024 - Current
University of Wisconsin-Madison
Madison - WI, USA

- 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**.
  - Tech stack: LangChain, LangGraph, LangSmith, PostgreSQL, PGVector, FastAPI, Slack SDK, Docker, Git, UV.
- 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.
  - See the application showcase and cost savings simulator.
- 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.

Data Scientist / ML Engineer
Jan 2023 - Jul 2024
Schlumberger | World's largest offshore drilling company
Remote - USA

- 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)
- 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.

Data Scientist / ML Engineer
May 2022 - Dec 2022
AB InBev | World's largest beer brewer company
Remote - Brazil

- Advanced **forecasting** modeling for multiple product categories with hierarchical reconciliation.
- Built a **scalable data pipeline** for automating pricing and promotion policies on big data using PySpark.

Data Scientist
Oct 2021 - Apr 2022
Remessa Online | Fintech
Remote - Brazil

- Exploratory data analysis, **time series forecasting**, and modeling **customer retention**.
- Engaged with **CRM** and **marketing** stakeholders to present easy-to-understand statistical reports.

CONSULTING EXPERIENCE

Data Science Writer
Apr 2024 - Current
Provided services for TrainInData
Remote - Germany

- Write SEO-optimized articles on data science [link]
- Contributed to User Guide pages and fixes for the feature-engine Python package.

Data Science Professor
Oct 2021 - Mar 2025
Provided services for DNC Group (Brazilian EdTech)
Remote - Brazil

- Recorded classes about data wrangling, statistics, recommender systems, clustering, model deployment, etc.
- Crafted materials from data wrangling to model deployment.

EDUCATION

M.B.A in Data Science & Analytics   May 2021 - Aug 2023
Universidade de São Paulo | Consistently ranked as the best university in Latin America   São Paulo - SP, Brazil
- **Grade:** 10 | **Dissertation:** A hybrid approach for predicting and preventing churn [code]

Ph.D. in Bioinformatics   Jul 2016 - Apr 2021
Universidade de São Paulo | Consistently ranked as the best university in Latin America   São Paulo - SP, Brazil
- **Grade:** 10 | **Thesis:** Identifying natural selection in Native American populations [code]

TECHNICAL SKILLS

Agentic AI & LLMs: LangChain, LangGraph, LangSmith, multi-agent workflows, RAG pipelines, prompt engineering, model fine-tuning

AI Engineering: Python, SQL, PyTorch, scikit-learn, HuggingFace, FastAPI, vector databases (PGVector), NLP, computer vision

MLOps & Cloud: Docker, GitHub Actions, CI/CD, MLflow, Weights & Biases, AWS, GCP, Azure

Data & Platforms: PostgreSQL, BigQuery, MySQL, PySpark, Databricks, Dataiku, geospatial data processing

Business & Communication: Data visualization, statistical modeling, A/B testing, explainable AI, stakeholder engagement

SELECTED PUBLICATIONS

- **Couto-Silva, C. M.**, et al. (2024). Mitigating Nonproductive Time: A Novel Algorithm for Dsl Fault Detection.
  OnePetro. DOI: 10.2523/IPTC-24515-MS.
- **Couto-Silva, C. M.**, et al. (2023). Indigenous people from Amazon show genetic signatures of pathogen-driven selection.
  Science Advances, 9(10). DOI: 10.1126/sciadv.abo0234.

SELECTED TALKS

Scipy 2025   Jul 2025 (scheduled)
Building an AI Agent for Natural Language to SQL Query Execution on Live Databases   Tacoma - WA, USA
- 4-hours tutorial | Live coding | See the schedule here!

PyData Global 2024   Dec 2024
PyTorch Workflow Mastery: A Guide to Track and Optimize Model Performance   Remote, Global
- 90-minute tutorial | Live coding | Code Repository

Nubank Meetup 2024   Oct 2024
Data Science Applied to Multiple Sectors: Challenges, Solutions, and Lessons Learned   Remote, Brazil

ADDITIONAL INFORMATION

LANGUAGES
- Portuguese (native) - English (professional / C1) - Spanish (intermediate)

SELECTED AWARDS
- **Best academic performance** award (2013)
- Highlighted didactic **web app** for clustering on **Streamlit Weekly Roundup**

OPEN-SOURCE CONTRIBUTION
- Contributed to **Feature-engine**, an open-source Python package implementing various feature engineering techniques.
- Developed **ProjectLens**, an open-source Python package that generates a well-structured text file with comprehensive project snapshots, optimizing context for LLM analysis and code review.

INTERNATIONAL EXPERIENCE
- Postdoctoral researcher position in Wisconsin, USA.
- PhD's internship in Barcelona - Spain.
- Master's internship in Buenos Aires - Argentina.
- Attended multiple conferences in South America, the USA, and Spain.