

Caio Seda Bittencourt

[Mail](#) - [Website](#) - [Github](#) - [Linkedin](#)

About Me

Master's student at PUC-Rio focused on Artificial Intelligence and Machine Learning. My research explores diffusion-based generative models for medical image segmentation in collaboration with INCA (Brazil's National Cancer Institute).

I have nearly six years of experience applying data science and deep learning to real-world problems in the financial sector, including credit risk modeling, causal inference, and automated analysis pipelines.

Recently, I've been expanding my research interests toward human-centered generative AI, exploring how these models can enhance creative and expressive applications.

Education

Ago, 2024 - Ago, 2026 | **MSc in Computer Science**, *PUC-Rio*

Field of study: AI/ML.

- Research on diffusion models for medical image segmentation.
- Collaboration with INCA (National Institute of Cancer), integrating computational research and clinical applications.

2017 - 2023 | **BSc in Computer Science**, *CEFET-RJ*

- Final paper: Causal study on credit policy changes during the COVID-19 pandemic using TMLE.
- Supported master's projects using Deep Learning and NLP for hierarchical text classification (TCE-RJ project).

Professional Experience

Ago, 2025 - Now | **Research Fellow**, *Tecgraf / PUC-Rio*

- Developed AI-based tools for medical image analysis in Videofluoroscopic Swallow Studies (VFSS), aiming to support dysphagia assessment through machine learning solutions.
- Generated synthetic medical images and performed data augmentation using StyleGAN2 to evaluate whether synthetic data could improve downstream discriminative model performance.
- Trained and evaluated diffusion-based models, including MedSegDiff and SDSeg, for supervised segmentation tasks, investigating whether diffusion models can compete with state-of-the-art discriminative approaches
- Designed and implemented a medical image annotation software with AI integration capabilities, enabling interactive labeling workflows and incorporation of trained models into the annotation process.
- Co-led the technical coordination of the research group by organizing labeling workflows, managing student contributions, and enforcing reproducibility and scalability across the software stack.

Technologies: Python, PyTorch, Docker, PySide

2021 - Ago, 2025 | **Data Scientist**, *CashU*

- Led causal impact analyses using Double Machine Learning, Difference-in-Differences, and TMLE to measure product effectiveness on client retention, credit utilization, and average ticket size, directly supporting strategic decisions.
- Developed a prospect analysis pipeline to assess credit risk for potential clients by evaluating behavioral model performance on client portfolios, reducing analysis time from one day to minutes, and enabling risk-adjusted contract pricing.
- Developed credit risk models for clients with and without transactional history, supporting the allocation of tens of millions of BRL in recurring monthly credit to previously unbanked clients.

- Rebuilt the feature engineering pipeline using DBT, enabling scalable, reproducible feature computation and allowing multiple daily executions to support near real-time risk monitoring.
- Built analytical dashboards for continuous monitoring of credit risk, product recommendations, and core company KPIs, improving visibility and decision-making across data, risk, and commercial teams.
- Conducted an observational causal study using TMLE to evaluate COVID-era credit policies, showing that clients accessing special credit lines exhibited lower default risk compared to traditional credit products.

Technologies: Python, SQL, DBT, MLflow, scikit-learn, XGBoost, Metabase, PostgreSQL

2019 - 2021 | **Data Scientist Intern, BNDES**

- Built web scraping and data ingestion pipelines in Python to collect public opinion data from Twitter and other social platforms, supporting a transparency initiative to assess public perception of the institution.
- Applied NLP techniques for sentiment analysis and topic exploration, producing analytical notebooks and visual summaries that informed marketing and communication strategies.
- Contributed to a federal highway traffic forecasting project by evaluating model performance metrics and assessing the usefulness of external data sources for demand prediction.
- Assisted in the development and evaluation of a fraud detection model for corporate credit card transactions, supporting risk mitigation efforts.
- Developed and maintained automated data pipelines and feature engineering workflows, including contributions to a feature store built on national labor market data (RAIS) to support analytical and modeling use cases.

Technologies: Python, PySpark, SQL, scikit-learn, PostgreSQL, HiveQL, Shell Scripting

Awards

2025 | **Nemesis** - AI for Public Expenditure Transparency (TCE-RJ)

- Winner of 1st Place INOVA Telebrasil Award

- Helped automate fraud detection and improve transparency in public oversight.
-

Languages

English: Advanced Intermediate (C1)

Portuguese: Native Language

Skills

Programming Language: Python, PostgreSQL, C++, C, Shell Scripting

Frameworks: PyTorch, Scikit-Learn, OpenCV, PIL, MLFlow, PySpark, XGBoost, PySide

Tools: Docker, DBT, Metabase, Linux