

Ruben Esteche

São Paulo, Brazil | rubenesteche@hotmail.com | +55 (085) 991125595 | linkedin.com/in/rubenesteche/ | github.com/ REsteche

SUMMARY

Passionate Data Scientist and AI/ML/MLOps Engineer specializing in Natural Language Processing and LLM architectures. Possesses a strong background in data science with over 6 years of professional experience, supplemented by backend development expertise. Proven track record in designing robust production architectures, building advanced conversational AI/RAG systems, and leading high-performing technical teams.

EXPERIENCE

Staff ML/AI Engineer — DataArt

July 2026 – Present | Remote

- Promoted to Staff role to oversee architectural strategy, technical standards, and cross-team execution for AI/ML systems serving major corporate financial institutions.

Senior ML/AI Engineer — DataArt

June 2025 – July 2026 | Remote

- Designed and maintained advanced dashboards and analytical data interfaces for corporate credit analysts utilizing graph and vector databases across preprocessing, ingestion, and agentic layers.
- Built metadata-extraction pipelines and advanced hierarchical chunking systems integrated natively with AWS Lambda ingestion workflows.
- Developed an AI agent transforming institutional databases into conversational applications using state-of-the-art RAG techniques including chain-of-thought structured responses, source tracking, and interactive financial plotting.
- Partnered with DevOps teams to build scalable production pipelines within AWS using Infrastructure as Code (IaC) and CI/CD best practices.
- Steered the technical interview program, evaluating senior Data Science, ML, and MLOps candidates and providing detailed competency matrices to HR and stakeholders.

Senior ML/MLOps Engineer — Bosch

July 2024 – June 2025 | Campinas, SP, Brazil

- Architected and developed production conversational AI solutions, enterprise chatbots, and Retrieval-Augmented Generation (RAG) platforms utilizing LLMs and deep NLP techniques at Bosch's AI Hub.
- Engineered scalable cloud production environments on Microsoft Azure using CI/CD pipelines, Docker containerization, and Kubernetes orchestration, reducing development lifecycle time by nearly 50%.
- Collaborated on computer vision deep learning product design to broaden the enterprise automation portfolio.

Lead Data Scientist & MLOps Engineer — NTT Data

March 2024 – July 2024 | São Paulo, Brazil

- Led the project AI division, managing backlog grooming, agile development sprints, team performance, and resolution of complex algorithmic bottlenecks.
- Acted as principal stakeholder and technical focal point ensuring absolute uptime and continuous support for live production AI functionalities.

Data Scientist & MLOps Engineer — NTT Data

January 2023 – February 2024 | São Paulo, Brazil

- Developed an enterprise NLP/ML chatbot system, achieving 92% classification precision and driving significant operational hours reduction for corporate clients.
- Deployed operational DataOps pipelines for automated model deliveries, establishing optimized developer workflows across teams.
- Governed end-to-end Data and MLOps architecture compliance throughout product release stages.
- Built generative AI-as-a-Service (AlaaS) solutions using structured microservices architectures.

ML Researcher & Backend Developer — Dell Lead

June 2022 – January 2023 | Ceará, Brazil

- Executed advanced web scraping, text classification, and custom ML on a global-scale accessibility initiative (Website Accessibility Layer).
- Integrated core NLP modules into high-throughput RESTful microservice APIs.
- Optimized model performance, improving training/inference speeds by 10%, compute-efficiency by over 15%, and accuracy parameters by 8%.

Data Scientist — Sugestão Consultoria

January 2020 – December 2021 | Ceará, Brazil

- Led technical design of data intelligence initiatives, interfacing with key corporate leadership to transform specific operational vulnerabilities into analytical solutions.
- Engineered predictive architectures satisfying a 90% floor accuracy requirement, prompting a 10% bottom-line expansion for consultancy clients.

Junior Data Scientist — Sugestão Consultoria

January 2019 – December 2019 | Ceará, Brazil

- Engineered early behavioral models reaching 95% forecast accuracy for demand planning and consumer trends.
- Automated continuous ingestion pipelines, slicing client data acquisition footprints by 25%.

EDUCATION

M.S. in Theoretical Quantum Physics (Minor in Computational Physics)	University of Pernambuco 2023
Postgraduate Degree in Data Science	Descomplica University 2022
B.S. in Physics (Minor in Physics)	Federal University of Pernambuco 2020

CERTIFICATIONS

Professional Machine Learning Engineer	— Google Cloud (Aug 2025 ID: 799bc86f891d4fc08d64c7aaca35841f)
Build and Deploy Machine Learning Solutions on Vertex AI	— Google Cloud (Jul 2025 ID: 17143663)
MLOps Introduction on GCP	— Google Cloud (Jul 2025 ID: 17141669)
Introduction to AI and Machine Learning on Google Cloud	— Google Cloud (Jul 2025 ID: 17143767)
Generative AI Fundamentals	— Google Cloud Skills Boost (2023)
School of Research Data Science	— ICTP - SAIFR (2022)
IBM Databases and SQL for Data Science	— IBM (2022)
Convolutional Neural Network with TensorFlow Design	— Federal University of Ceará (2020)

TECHNICAL SKILLS

Data Science:	NLP, LLMs, Generative AI, RAG, Transformers, Deep Learning, Computer Vision, CNNs, Statistical Modeling
MLOps/Data:	Docker, Kubernetes, CI/CD pipelines, Azure DevOps, DataOps, Vertex AI, Vector & Graph
Databases Cloud:	Google Cloud Platform (GCP), AWS (Lambda, ECS, S3, API Gateway), Microsoft Azure
Backend/Lang:	Python, FastAPI, Flask, Django, SQL, C, C++, NoSQL (Redis, MongoDB, Elasticsearch)
Languages:	English, Portuguese, Spanish, French