Bruno Stefoni

New York, NY | brunostefoni12@gmail.com | +1 646 203 7124 | **in** LinkedIn | **O** GitHub | **🎓** Scholar

PROFESSIONAL SUMMARY

Seven years of experience with AI, Machine Learning and Data: leading projects to deploy AI and ML solutions to the cloud; building data pipelines to train models; designing A/B tests and tracking KPIs.

EXPERIENCE

•CloudWerx   Nov 2023 – Present
Senior AI / ML Engineer   Manhattan, New York

- Leading multiple customer projects, deploying ML and AI solutions using GCP, Vertex AI, Vertex Search, Agent Builder, Big Query ML, Kubeflow, LLMs, Stable Diffusion, etc.
- Example projects: custom OSS LLM RAG chatbot implementation, Augmented Search using RAG and Gemini, fine-tuning pipeline for diffusion model, MLOps Kubeflow pipeline for model CI/CD.
- Example customer industries: cyber security, ecommerce, finance, design, advertising, robotics, real estate, data.

•Aetna - CVS Health   Sep 2023 – Oct 2023
Data Scientist   Manhattan, New York

- Designed A/B test experiment with 9 MM MAU App to measure campaign impact on customer adoption KPIs.
- Analyzed data in GCP Big Query to perform demographic analysis for a live A/A experiment.

•Sortile   Jun 2023 – Aug 2023
Data Scientist   Manhattan, New York

- Built a 92% accurate XGBoost classifier to identify black/white fibers using light wavelength sensor data.
- Developed and benchmarked XGBoost, LGBM and Deep Learning models. Deployed XGB model on AWS Lambda.

•Google   Jun 2022 – Aug 2022
Data Scientist Intern   Mountain View, California

- Conducted detailed market research and graph analysis by building a network dataset from diverse data sources using PLX, BigQuery and Python.
- Detected up to 15% of user churn by continuous survival and time series analysis of Software adoption rate data.
- Presented insights to internal stakeholders and management. Proposed a new methodology for analysts.

•STP   May 2018 – May 2022
Data Scientist   Santiago, Chile

- Built and implemented a company-wide chat bot on prem cloud server to support management and on-site workers.
- Drove STP's revenue by identifying routes with up to 25% difference between planned and real bus speeds.
- Prevented $500K investment on less accurate tech by launching a statistical study.
- Detected data outliers due to faulty batteries in 20% of STP's electric fleet via statistical analysis.
- Increased main contract KPI by 6% by scraping Twitter data to alert service disruptions.

PROJECTS AND PUBLICATIONS

•Anky Developed RAG + SFT/DPO fine tuned LLM with llama.cpp using social network data

•S&P Global Capstone Project News classification by fine-tuning BERT from unlabeled data (97% accuracy)

•Master's Thesis Publications Two papers published on computer microscopic simulation

•Deep Learning Competition Next frame video prediction using Transformer and CNN (ResNet) in JEPA network

•Text Topic Analysis LDA and GloVe word embeddings on large academic thesis corpus

•Twitter Sentiment Analysis Using fastText and fine-tuning BERT for content strategy assessment

•Hackathon Winner Led team of 6, designed and developed an app for transit users

EDUCATION

•NEW YORK UNIVERSITY   2021 – 2023
Master of Science in Data Science   GPA: 3.8/4.0

- Coursework: Deep Learning, Machine Learning, Big Data, Optimization & Linear Algebra, Probability & Statistics
- TA of Data Science Class at NYU Business School for two semesters

.

•UNIVERSIDAD DE CHILE 2018
Bachelor and Master's degree in Engineering Final grade 6.6/7.0

– Graduated with highest distinction in top 10% GPA of cohort
– Granted Outstanding Student Award
– TA of five different courses, e.g., Stochastic Processes, Network Optimization
– Key Modules: Data Mining, Computer Science I and II, Probability, Statistics, Optimization, Operations Research

Skills

Certifications: Google Professional Machine Learning Engineer
Programming Languages: Python, R, MATLAB, Java
AI Libraries: Transformers, Langchain, OpenAI API, Anthropic API, Instructor, Pydantic, HuggingFace
ML Libraries: PyTorch, PyTorch-Lightning, XGBoost, LGBM, scikit-learn, statsmodels
Cloud Platforms: Google Cloud Platform (GCP), AWS
Databases: SQL (SQlite), PostgreSQL (AlloyDB), BigQuery, Vector DBs (Milvus, Pinecone)
DevOps: Git, Docker, Terraform, Bash, Linux
Data Analysis: Pandas, NumPy, Plotly, Seaborn, Matplotlib
NLP: NLTK, SpaCy, fastText
Big Data: pySpark, Dask, Dataflow, Apache Beam
Languages: Native English and Spanish speaker, Intermediate Italian