

MARCELO PRATES, PHD

Data Scientist & Machine Learning Engineer

@ marceloop@gmail.com in marceloprates marceloprates



EXPERIENCE

Large Language Models Consultant

Vortigo

Nov 2023 – Now Porto Alegre, Brazil (Remote)

- As LLM consultant, I collaborated with a Brazilian tech company in the design and implementation of assistant ChatBots informed by proprietary source code and spreadsheet knowledge bases by leveraging OpenAI's paid API and pretrained Large Language Models such as GPT3.5 and GPT4.
- Tech stack:
 - Programming Languages: Python
 - ML & Analytics: Pytorch & Pytorch Lightning, Pandas, Scikit-learn, Numpy, Jupyter Notebooks, AWS Sagemaker
 - NLP & LLMs: OpenAI's API, HuggingFace models & Transformers library, LangChain, BertTopic

Senior Data Scientist

Dataside

Oct 2023 – Now São José dos Campos, Brazil (Remote)

- As a Machine Learning, Natural Language Processing and Computer Vision consultant, I collaborated with a Brazilian tech company in the design and implementation of assistant ChatBots informed by proprietary knowledge bases by leveraging OpenAI's paid API and pretrained Large Language Models such as GPT3.5 and GPT4. I also collaborated in the development of computer vision solutions for photogrammetry-based liquid volume estimation.
- Tech stack:
 - Programming Languages: Python
 - ML & Analytics: Pytorch & Pytorch Lightning, Pandas, Scikit-learn, Numpy, Jupyter Notebooks, AWS Sagemaker
 - NLP & LLMs: OpenAI's API, HuggingFace models & Transformers library, LangChain, BertTopic
 - Deployment: Flask, Docker
 - Computer Vision: OpenCV, Scikit-Image, Open3D, Pytorch & Numpy (for semantic segmentation), Shapely

Sabbatical Period

To focus on my generative art projects

Apr 2023 – Oct 2023 Porto Alegre, Brazil

- I took a short sabbatical period to focus on my generative art projects and maintain / improve existing Python packages I had built to help me in my artistic process, including prettymaps

Machine Learning / Computer Vision Consultant

ConstructIN

Mar 2022 – Apr 2023 Porto Alegre, Brazil (Remote)

- As a Machine Learning & Computer Vision consultant, I helped ConstructIN design computer vision solutions for automating construction site monitoring through 360-degree photography with state-of-art in machine learning technology.
- In addition to research and development efforts, I led a team of 3 data scientists responsible for the development of 4 CV applications for construction site monitoring, the main one being a dashboard clients could use to assess the speed of the construction process, which was powered by a semantic segmentation module built by our team.
- Tech stack:
 - Data warehouse & ETL: AWS

SKILLS

- Machine Learning Data Science
- Operations Research
- Combinatorial Optimization
- Computer Vision LLMs & NLP
- Geospatial programming
- Graph Deep Learning Management
- ML for Health


LANGUAGES & LIBS

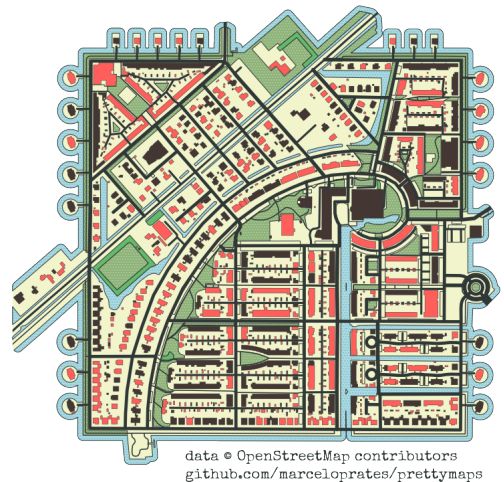
- Python ●●●●●●●●
- LLMs, Langchain, Huggingface ●●●●●●●●
- Pandas, Sklearn, Numpy ●●●●●●●●
- Google OR Tools ●●●●●●●●
- OpenCV, Open3D, Skimage ●●●●●●●●
- Django, Flask ●●●●●●●●
- Tensorflow ●●●●●●●●
- Pytorch ●●●●●●●●
- C/C++ ●●●●●●●●

PROFICIENCY


- Portuguese (native speaker) ●●●●●●●●
- English ●●●●●●●●
- Spanish ●●●●●●●●

MOST PROUD OF

-  **Prettymaps**
Python library for drawing pretty maps from OpenStreetMap data (1st place on Hacker News, >10.000 stars, ranked among the 2000 most starred repositories on GitHub)



data © OpenStreetMap contributors
github.com/marceloprates/prettymaps

-  **Gender Bias in Machine Translation**
Our 2018 research measured the prevalence of male defaults in Machine Translation tools

2. Programming Languages: Python
3. ML & Analytics: Tensorflow & Keras, Pytorch & Pytorch Lightning, Pandas, Scikit-learn, Scipy, Numpy, Matplotlib, Jupyter Notebooks, AWS Sagemaker
4. Computer vision: Open3D, OpenCV, Scikit-image
5. Deployment: Flask, Django, Docker

Generative Art Teacher

Responsive Cities

📅 Nov 2022

📍 Porto Alegre, Brazil

- Taught a course on generative art history & principles and on useful tools and libraries for creative coding

Principal Data Scientist

Condati

📅 Nov 2021 – Dec 2023

📍 Menlo Park, California (Remote)

- My main responsibility as a Principal Data Scientist at ConDati was to improve an existing Machine Learning solution meant for optimizing bidding strategies for digital marketing campaigns. The financial return generated by the existing solution was performing below the goal for quite some time when I started at ConDati, but our team was able to diagnose problems and, in the time frame of less than one year, bring the model's performance back to the target defined with the client.
- Tech stack:
 1. Data: mySQL
 2. Programming Languages: Julia & Python
 3. ML & Analytics: Tensorflow.jl, Torch.jl, MLJ.jl, Flux.jl, Scipy, Numpy, Pandas, Pytorch, Jupyter Notebooks, AWS Sagemaker

AI Researcher & Project Leader - ML for Health

Samsung Research Brazil

📅 Mar 2020 – Nov 2021

📍 Campinas, Brazil (Remote)

- I was initially hired as an AI Researcher to work on a project consisting in the development of an ML-powered solution for real-time health monitoring on wearable devices. I worked in several areas, including: 1) the design of data collection protocols; 2) researching the state-of-the-art in AI for health monitoring, the prototyping, development and testing of ML solutions for health monitoring on wearable devices; 3) and the implementation of a memory-constrained solution written in C to run on the Samsung Galaxy Watch 4.
- I started leading the research, development and data team from September 2020 to November 2021. During my period as team leader, we were able to present convincing results of the effectiveness of our solution to the Samsung headquarters, and the project was successfully incorporated to the Galaxy Watch line worldwide (starting at model 4) as a ML-powered health-monitoring app.
- Tech stack:
 1. Programming Languages: Python, Julia, C/C++
 2. ML & Analytics: Tensorflow & Keras, Pytorch & Pytorch Lightning, Pandas, Scipy, Numpy, Scikit-learn, Matplotlib, MLJ.jl, Flux.jl, Jupyter Notebooks, AWS Sagemaker
 3. Deployment: C, ONNX, custom Python-to-C transpilers built by our team, Samsung's Tizen OS

Data Scientist

Poatek IT Consulting

📅 Jun 2019 – Mar 2020

📍 Porto Alegre, Brazil

- Worked on a wide range of interesting projects, including 1) Exact and heuristic methods for solving combinatorial optimization problems, mostly related to vehicle routing 2) Computer vision and NLP applied to the problem of cataloging documents and extracting & analyzing data from them; 3) NLP applied to the problem of Named Entity Recognition (NER) and sentiment analysis; 4) Credit risk modelling; 5) Relationship extraction and visualization applied to geospatial data. Most of these projects included machine learning or deep learning to some extent (convolutional neural networks, gradient boosting / ensemble methods and pre-trained language models).
 - Tech stack:
 1. Programming Languages: Python, Julia, C/C++
 2. ML & Analytics: Tensorflow & Keras, Pytorch & Pytorch Lightning, Pandas, GeoPandas, OpenCV, Scikit-image, Scikit-learn, Scipy, Numpy, Matplotlib, Jupyter Notebooks,
 3. Combinatorial Optimization: JuMP, Google OR-Tools
 4. Deployment: Flask, Django, Docker
-

PhD in Computer Science

Federal University of Rio Grande do Sul (UFRGS)

📅 Aug 2015 – Jul 2019

📍 Porto Alegre, Brazil

- **AI Ethics:** On Quantifying and Understanding the Role of Ethics in AI Research (GCAI 2018), Assessing Gender Bias in Machine Translation (NCA 2020)
 - **Graph Deep Learning & Combinatorial Optimization:** Learning to Solve NP-Complete Problems (AAAI 2019), Graph neural networks meet neural-symbolic computing (IJCAI 2021)
-

Figures created by me. Resume was created using AltaCV class by LianTze Lim (liantze@gmail.com).