

Professional Summary

Machine Learning Engineer focused on the end-to-end training and deployment of AI models. I combine a strong foundation in traditional Deep Learning and Computer Vision with advanced expertise in LLM fine-tuning and alignment. My work centers on training models that are not only accurate but also cost-effective and secure, utilizing techniques like quantization and distributed training. In addition to custom modeling, I design production-grade systems that integrate models from OpenAI, Google, and Anthropic into scalable backend infrastructures. I bring a research-driven approach to solving real-world problems through custom-trained AI architectures, bridging my academic background with practical industry experience. With native bilingual fluency and multicultural experience from living in both Brazil and the United States, I offer strong communication skills and adaptability to diverse, global teams.

Experience

Cadastra (Contracted to Grupo Boticário)

📅 Apr/2025 – Mar/2026
Brazil

Machine Learning Engineer

- Maintains and updates a GenAI Gateway system connecting multiple third-party LLM providers like OpenAI, Google, and Anthropic via Azure and VertexAI.
- Engineered a feature for the company's GenAI SDK using LiteLLM to allow for model-switching and unified classes across frameworks like Agno, CrewAI and Google ADK.
- Developed APIs for image/audio transcription and RAG-as-a-service using FastAPI, LangChain, and LLMs.
- Built an autonomous system on GCP that reduced a manual 1-hour data update process to just 5 minutes.
- Used various RAG methods, like LLM Embeddings with third-party models, and maintenance of a Cache-Augmented Generation (CAG) implementation.
- Performed database mapping and complex queries using PostgreSQL and SQLAlchemy.
- Implemented Unit Testing and automated evaluation pipelines using LLM-as-a-judge to ensure code and prompt quality.
- Used Kanban methods and metrics for organizing tasks and measuring the team's performance.

Docket

📅 Oct/2024 – Nov/2024
Brazil

Machine Learning Engineer

- Fine-tuned multimodal models for Named Entity Recognition (NER) on legal deeds to assist in property ownership data extraction.
- Used Google Cloud Vision API to extract text from high-variance notary documents.
- Managed the pre-processing and curation of text/image datasets stored in AWS S3.

Weni

📅 Nov/2023 – Oct/2024
Brazil

Machine Learning Engineer

- Fine-tuned models up to 70B parameters using PyTorch, DeepSpeed, Hugging Face Accelerate, and PyTorch FSDP.
- Aligned LLMs using SFT and RLHF/DPO for Zero-shot classification and Q&A tasks.
- Applied AWQ Quantization to 70B models to make production deployment more efficient.
- Compared RAG pipelines using LangChain, Hybrid Search and Cohere Reranking logic with Amazon Bedrock.
- Developed semi-autonomous training pipelines to move experiments from Google Colab to RunPod.io, enabling multi-GPU scaling and increasing the team's MLOps maturity.

Bio-inspired Computer and Engineering Lab (LABCIN)

📅 Aug/2020 – Oct/2024
Brazil

Researcher

- I was the Lead Author of research on [Transformers and Few-Shot Learning applied to digital pathology and clinical skin images](#), published in the IEEE Journal: Expert Systems with Applications (2023).
- Investigated Variational Autoencoders (VAEs) and Few-Shot Learning for medical feature extraction.
- Peer-reviewed for the IEEE Journal of Biomedical and Health Informatics (ORCID: 0000-0003-1057-2785).

- Participated in research groups for MLOps and Explainable AI (XAI).
- Developed a hybrid OCR pipeline using OpenCV and Tesseract, and used YOLO for element detection.
- Developed NLP and Named Entity Recognition (NER) pipelines using Hugging Face to process and extract data from unstructured text.
- Curated Python-based Docker images to ensure they were vulnerability-free or low risk.

VERT Transformação em TI
Data Scientist (Intern to Full-time)

📅 Jun/2021 – Jun/2022
Brazil

- Led client meetings and developed risk analysis frameworks for government bidding processes (State of Pará) and DETRAN-ES financial projects throughout both internship and full-time roles.
- Architected ETL processes in SAS (9.4/Guide) and managed database queries and mapping with PostgreSQL and SQLAlchemy.
- Developed web scrapers using BeautifulSoup and Selenium for automated public data collection.

Education

Federal University of Espirito Santo (UFES)
Bachelor's Degree in Computer Science

📅 Apr/2019 – Expected Mar/2026
Vitoria, ES, Brazil

- Coursework completed in 2024. Emphasis on Machine Learning and Software Engineering.

Skills and Expertise

- **Generative AI:** LLM Orchestration, Agentic AI (Agno, CrewAI, Google ADK, LangGraph), LangChain, RAG, DPO, RLHF, SFT, AWQ, Embeddings.
- **Core ML/DL:** PyTorch, TensorFlow, Keras, CNNs, GANs, YOLO, Transformers, Scikit-learn, Pandas.
- **Engineering:** FastAPI, PostgreSQL, SQLAlchemy, Unit Testing, Multi-GPU Scaling (DeepSpeed, FSDP, Accelerate).
- **Infrastructure:** Docker, Vulnerability Management, GCP, AWS Sagemaker, Weights & Biases, GitLab CI/CD.
- **Languages:** Native English, Native Portuguese, Beginner French.

Volunteering

Introcomp - Introdução à Computação (Introduction to Computing)
Teacher

📅 Aug/2019 – Nov/2022
Vitoria, ES, Brazil

- Taught Python programming to public high school students and created Jupyter notebooks for use as study materials and teaching guides.
- Lead the data collection and class streaming committees.
- Co-authored an article on programming teaching challenges for public school systems for the 2021 Brazilian Symposium on Computing Education.

Dermatological Assistance Program (PAD-UFES)
Information Technology Support Technician

📅 Sep/2021 – Dec/2021
Vitoria, ES, Brazil

- Provided IT support for a rural healthcare program, assisting with patient registration systems and data collection for skin cancer datasets.

**Laboratório de Administração de Redes
(Network Administration Laboratory) - UFES**
Laboratory Assistant

📅 Apr/2019 – Mar/2020
Vitoria, ES, Brazil

- Assisted professors with exam applications and managed network and user account issues within the Computer Science labs.