Leonardo Azzi Martins

Email: lamartins@inf.ufrgs.br
LinkedIn: linkedin.com/in/leonardoazzi
Mobile: +55 51 99692-1684
Github: github.com/leonardoazzi

Professional Summary

Machine Learning Scientist with 3+ years of experience researching, designing and experimenting ML and LLM-powered systems for enterprise products and services. Expertise in NLP, LLM fine-tuning, model evaluation and data-centric AI. Skilled in PyTorch, Hugging Face Transformers, MLFlow, SQL, and cloud platforms (Azure, Databricks). Demonstrated ability to deliver AI features with measurable business impact through collaboration with research, engineering and product teams.

Skills Summary

- **LLMs & NLP**: Hugging Face Transformers, LLM fine-tuning, prompt engineering, LoRA PEFT, RAG pipelines, grounding techniques, knowledge distillation, LLM evaluation, LlamaIndex, Unsloth, SpaCy.
- **ML & Deep Learning**: PyTorch, TensorFlow, Scikit-Learn, Pandas, Numpy, data science and enginering, experiment tracking, SHAP, MLFlow, model monitoring.
- **Cloud & Data**: Azure, Databricks, PostgreSQL, Linux.
- **Programming**: Python, SQL, C/C++, Typescript, Flask, Git.
- **Soft Skills**: Team collaboration, experimentation, technical writing, public speaking, collective leadership.

Experience

- **DB, Innovation Lab - R&D Team** — Remote
  Machine Learning Research Engineer — *May 2020 - current*
  - **R&D for New Product Development**: Conducted research projects to support new products and services with AI, involving Fine-Tuning, Agent Evaluation, Prompt Engineering, LLM Observability and Computer Vision.
  - **Tech Lead**: Reference in AI / ML technical and strategical decisions. Conducted teams of developers, intern mentorships, agile cerimonies, project planning and business proposals.
  - **LLM Engineering, TCE-RS**: Proof-of-Concept Prompt Engineering experiments, tracked with MLFlow, for productivity in accounts court processes. Resulted in new features in production for TCE-RS audit software.
  - **Vision Detector, Edenred TicketLog**: Proof-of-concept object detection model to recognize digits in Brazilian car odometers, achieving significantly higher accuracy than baseline OCR tools

- **DB, Business Unit** — Porto Alegre, Brazil
  Software Developer — *Mar 2018 - Jan 2020*
  - **QA Consultant**: Delivered QA consulting solutions for clients in the Toronto/Ontario region, improving software reliability through automated test strategies.
  - **Test Automation Engineer, Avionics**: Implemented test automation for avionics embedded systems using Python and Lua, ensuring compliance with mission-critical safety standards.
  - **Internship, Avionics**: Worked on software development, data modeling and test maintenance during internship.

Highlighted Projects

- **Pipeline for Fine-tuning Expert Language Models with Private Data (2025)**: The project aims to improve the process of efficiently fine-tuning language models using private data for specific business objectives. It proposes a development pipeline as an alternative to closed, general-purpose API-based models. The research investigates current fine-tuning methods, their use cases, and their trade-offs in cost and quality. Finally, it seeks a decision-making process to guide the selection of fine-tuning approaches that minimize risks and ensure solution quality.
  - **Skills**: SFT, PEFT, Reinforcement Learning, RLHF, Knowledge Distillation, SLMs

- **Fine-tuned Misinformation Sequence Classifiers on Language Models (2025)**: Investigates and compares the effectiveness of two language model approaches for misinformation detection in Brazilian Portuguese, based on NLU and NLG. Fine-tuning experiments were conducted with BERTimbau and Qwen3-0.6B, and their performance was compared with classic ML models and feature embeddings. While BERTimbau outperformed previous approaches with an F1 score of 0.857, Qwen showed weaker performance at 0.787, only slightly above the original SVM baseline of 0.778.
  - **Skills**: Python, SFT, HF Transformers, PyTorch, Scikit, Pandas, Numpy, Imblearn, Seaborn

- **TCE-RS: AI experiments for productivity in Macroprocess (2025)**: Developed LLM-based PoCs for automating legal and audit text generation, including instruction reuse, abstractive summarization of audit findings, and chapter-level analytical summaries. Designed datasets, prompting strategies (one-shot/few-shot), and evaluation pipelines using similarity metrics and qualitative user feedback. Achieved competitive syntactic similarity scores and identified data and model scaling paths to improve robustness. Validated the feasibility of integrating LLMs into audit workflows to accelerate analysis and reduce manual drafting effort.
  - **Skills**: Prompt Engineering, LLM Evaluation, LLM-as-a-Judge, MLFlow, Azure OpenAI, Pandas

- **Evaluating Text-to-SQL agents for complex databases (2025)**: Led a research project on Text-to-SQL systems using LLMs, benchmarking the Analy PoC and commercial/genAI agents on the Spider 2.0 dataset. Investigated latency, correctness, and schema-linking challenges, identifying architectural bottlenecks and proposing state-of-the-art agentic and retrieval-based improvements. Adapted and evaluated frameworks such as SpiderAgent, ReAct-based agents, and Databricks Genie on real and synthetic databases. Delivered technical insights that guide future digital products and strengthen the organization's internal expertise in LLM-based data analysis.

- **Skills**: SQL, Agents, Agentic Architectures, Benchmarking

- **Comparing RAG and Fine-tuning for Question-Answering in private documents (2025)**: Designed and executed a comparative evaluation framework for LLM-based question-answering systems, benchmarking fine-tuned Llama 3.1 models against RAG pipelines using GPT-3.5 and GPT-4o. Built synthetic QA datasets, implemented MLFlow and Mosaic AI–based evaluation workflows, and assessed correctness, safety, and response quality with LLM-as-a-Judge. Findings indicate that the most effective approach combines RAG for knowledge retrieval with targeted fine-tuning, with future gains possible through RAFT-style fine-tuning. Delivered technical insights that guide MLOps practices and the adoption of reliable generative AI solutions.
  - **Skills**: RAG, LlamaIndex, SFT, Unsloth, Hugging Face PEFT, Databricks Agent Evaluation, MLFlow, Azure OpenAI, Python

- **Fine-Tuning for Python Unit Test Generation using PEFT and LoRA (2024)**: Fine-tuned a CodeGemma 7B LLM on Python functions with PyTest unit tests using PEFT and LoRA adapters. Conducted dataset curation, cleaning 15k+ high-quality instances and preprocessing for training. Designed and tested a fine-tuning pipeline on Databricks, addressing VRAM and distributed training constraints. Explored Transformer architecture and attention mechanisms to select optimal hyperparameters for task-specific adaptation.
  - **Skills**: SFT, Unsloth, Hugging Face PEFT, PyTest, PySpark DeepSpeed, Pandas, Python

- **Prediction of Academic Dropout at the Polytechnic Institute of Portalegre (2024)**: A Machine Learning model was developed to predict student dropout at IPP. After exploratory analysis and algorithm spot-checking, three pipelines were optimized and evaluated. The best-performing model, based on AdaBoost, has 0.756 F1-score. An interpretability analysis was conducted to explain model previsions and support educational policies.
  - **Skills**: Spot-checking, Interpretability, Model Evaluation, Sklearn, Optuna, Pandas, Numpy, Imblearn, Seaborn, SHAP

- **Data-centric AI for Vision (2024)**: Typical ML projects "freeze" the data and iterate over hyperparameters in an attempt to improve metrics. However, this paradigm often results in fragile models and architectures when deployed in real-world applications, as it propagates errors from noisy data. Data-Centric AI involves methods to systematically characterize, assess, and monitor the data used to train and validate models. Techniques were researched to apply Data-Centric AI within a pipeline for Computer Vision projects. These methods are evaluated through case studies involving real vision projects developed at DBLab.
  - **Skills**: Data Engineering, CRISP-DM, Cleanlab, YOLO, PyTorch, Python

- **3D Computer Vision for Volume Estimation (2024)**: This project aimed to evaluate and apply 3D computer vision methods for depth estimation in indoor environments. The methodology involved reviewing algorithms, datasets, and metrics, selecting techniques suitable for accurate 3D reconstruction, and demonstrating them through disparity maps using ORB, HSM, and Vis-MVSNet. The study mapped the strengths, limitations, and applicability of each approach for volume estimation.
  - **Skills**: 3D Vision Algorithms, Deep Learning Architectures

- **Vehicle Odometer Reading Using Computer Vision AI (2023)**: A proof-of-concept was developed to enable automatic reading of vehicle odometers using OCR and object detection techniques. After a literature review, an iterative experimental process validated state-of-the-art methods with EasyOCR and YOLOv8. OCR alone showed low accuracy, while a pipeline combining odometer detection and digit detection with YOLOv8 achieved satisfactory results. This study demonstrates the potential of object detection pipelines for reliable digit recognition in vehicle odometers.
  - **Skills**: YOLO, Pandas, Scikit, EasyOCR, Model Evaluation, Python

Education

- **Federal University of Rio Grande do Sul (UFRGS)** — Porto Alegre, Brazil
  BSc. Computer Science — *July 2021 - December 2027*
  Relevant courses: *Natural Language Processing, Machine Learning, Artificial Intelligence, Data Structures, Fundamentals of Databases*
  Attended events: *ERAMIA '25: 1st Southern Regional School on Machine Learning and Artificial Intelligence; SIBGRAPI '23: 36th Conference on Graphics, Patterns and Images*

- **Federal Institute of Education, Science and Technology (IFSUL)** — Charqueadas, Brazil
  Mechatronics Technician degree — *Mar 2014 - April 2018*

Certifications and Courses

- Reinforcement Learning: How to Teach Robots to Maximize Rewards in Practice - 1st ERAMIA-RS (2025)
- Interpretation of Machine Learning Models - 1st ERAMIA-RS (2025)
- Get Started with Databricks for Machine Learning - Databricks Partner Academy (2024)
- Python for Natural Language Processing - ICMC USP (2021)