

Diogo Antunes

📍 Braga, Portugal (Flexible) ✉ contact@diogo-antunes.dev ☎ +351 939 911 077 🔗 diogo-antunes.dev
 in diogo-antunes2 🌀 0xdbff 🌐 vaisys.com 🐙 git.vaisys.com (vpn)

About Me

I am a software engineer and AI researcher specialized in **Machine Learning** and **Data Science**. With a solid background in linear algebra, statistics, calculus, algorithms, and distributed computing, I am an engineer who goes beyond coding to understand the intricate complexities underlying machine learning. This ambition and broad perspective often drive innovation and result in impactful systems. I am a constant learner in this rapidly evolving field.

I am not only a deeply technical engineer, but also an entrepreneur who is always looking for new ideas and opportunities on how to bring value to society with the use of AI, bringing a clear, value-driven perspective to the use of safe Artificial Intelligence.

Experience

Defense Point — ML Engineer, Data Scientist

Braga, Portugal
Hybrid
November 2024 – **Present**

- Designed and Implemented ML solutions for advanced IAM (Identity and Access Management) with the help of clustering algorithms and customized neural networks.

Checkmarx — Machine Learning Engineer

Braga, Portugal
Hybrid
June 2024 – October 2024

- Solely built tooling to fine-tune open-source state-of-the-art autoregressive **LLMs** with **SFT** (Supervised Fine-Tuning) and integrated **RLHF** (Reinforcement Learning with Human Feedback), allowing for proprietary alignments of LLMs.
- Solely responsible for building the instruction dataset for SFT; also built partial tooling to transform internal raw information into relevant instructions for the model.

Checkmarx — Machine Learning Engineer (Internship)

Braga, Portugal
Hybrid
March 2024 – June 2024

- Solely designed and built internal tooling to fully manage the **lifecycle** of transformer models, from design to training (pretraining and finetuning), to automated finetuning, and production.
- Solely designed a transformer architecture for **sequence classification** tasks (encoder-only), which can be employed in multiple domains by leveraging new datasets, predictive classes (multi-class, multi-label or binary), and a potential new tokenizer.

Vaisys (Startup Founding) — Principal Engineer, AI Researcher

Braga, Portugal
On Site
January 2023 – **Present**

- Designed a complete system for price forecasting and technical analysis study, independent of the market/asset.
- Designed, trained, and integrated a complete **transformer neural network** (decoder-only) into the system.
 - The model, designed with 1B parameters, achieved 0.94 cross-entropy loss on a test set during the pre-training stage (trained on 10B tokens), meaning a 39% chance that the predicted token is correct, which is significant.
 - By leveraging confidence intervals and fine-tuning the model with a custom loss function, it has the potential to achieve significant forecasting accuracy. Currently it is being updated in the fine-tuning stage.
- Currently studying and researching **reinforcement learning**, using CARLA and simpler games to test the agent's potential under different environments.
- Currently designing an RL agent intended to execute financial market actions by combining confidence intervals extracted from the predictive model.

Education

- BS** **Polytechnic Institute of Cavado and Ave**, Software engineering 2021 – 2024
- Grade: 80%
 - Internship Thesis — Scan Optimization With AI [↗](#)
 - **Coursework:** Algorithms, Linear Algebra, Statistics, Calculus, Imperative Programming, Object-Oriented Programming, Computer Vision, Natural Language Processing, Artificial Intelligence and others.

Professional Certifications

- Nvidia — Fundamentals of Accelerated Computing with CUDA C/C++ [↗](#) June 29, 2022
- Nvidia — Fundamentals of Deep Learning [↗](#) July 6, 2022
- Nvidia — Fundamentals of Accelerated Data Science [↗](#) July 13, 2022
- Nvidia — Disaster Risk Monitoring Using Satellite Imagery [↗](#) February 10, 2023
- Google — Advanced Data Analytics [↗](#) June 28, 2023
- Checkmarx — Security Champion [↗](#) March 12, 2024
- LinkedIn — Certifications [↗](#)

Publications

- Scan Optimization With AI** July 19, 2024
- Diogo Antunes
- A Modular Sequence Classification approach
- Sequence Classification [↗](#)

Projects

- Housify, Complete Generative AI system** github.com/geniodynamics [↗](#)
- Developed an internal AI web server to perform distributed inference on open-source AI models such as LLaVa, Stable Diffusion, and Pix2Pix.
 - Built a FastAPI web server to connect with clients using JWT and standard endpoints.
 - Developed an Android mobile application to support client-side functional requirements.
 - Tools used: Python, PyTorch, Kotlin, Nginx, PostgreSQL, Redis.
- TuxTech** github.com/0xdbff/tuxTech [↗](#)
- Complete e-commerce system built with Django.
- Traffic Signal Detection** github.com/0xdbff/visao-por-computador [↗](#)
- Computer vision solution for traffic signal detection using OpenCV in C++.

Technologies

Languages: Python, C/C++ CUDA, TypeScript, Kotlin, Rust.

Technologies: Pytorch, Tranformers, scikit-learn, FastAPI, Django, PostgreSQL, Redis, Nginx, Docker and others.

AI-Domains: Deep Learning, Reinforcement Learning, Machine Learning, Computer Vision, Search Algorithms.

Skills: Large Language Models, Model Training, Model Quantization, Inference, Distributed Computing, Asynchronous Programming and others.