Gabriel Coelho

Braga, Portugal gabrielcoelho4398@gmail.com +351910803774 gabrielcoelhoo
gabrieljdcoelho

I am a data scientist with a strong academic and industry background, currently pursuing a Ph.D. in Information Systems and Technologies at University of Minho. My research focuses on Enhancing Privacy and Fairness via Synthetic Data Generation and Deep Learning, reflecting my interest in building responsible AI systems. At CEGID, I work on projects that directly impact the business, applying data science and machine learning to solve real-world problems and drive value. With previous experience in research and a passion for continuous learning, I bring an analytical, ambitious, and resilient mindset to every challenge.

Education

PhD in Information Systems and Technology *Feb 2024 – Present*
University of Minho

During the PhD, the focus has been on synthetic data generation models (such as Gans and diffusion models) and how these can be used to increase the privacy and fariness of data for training machine learning models.

- **Thesis Title:** Enhancing Privacy and Fairness via Synthetic Data Generation and Deep Learning
- **Advisor:** Paulo Cortez

MSc in Engineering and Management of Information Systems *Sep 2016 – Dec 2021*
University of Minho

A course with a strong software engineering and management component, where along the way I specialized in artificial intelligence and software development. My journey ended with my master's thesis being graded with the highest grade, 20.

- **Thesis Title:** Generation of Sensory Data from the Vehicle Interior for a Machine Learning Approach to Anomaly Detection
- **Advisor:** Paulo Cortez

Experience

Data Scientist *Apr 2025 – Present*
CEGID

- Part of the AI PODS team, working with the product to maintain and devise new ways of using artificial intelligence to improve it;
- Responsible for keeping 12 models in production, in a mature mlops lifecycle using Azure, always aiming to improve their metrics;
- Data scientist responsible for changing the strategy of CEGID's largest product from a more traditional approach to generative AI.

Data Scientist *Mar 2022 – Apr 2025*
Deus.AI

- Integration of LLMs into pipelines for tasks such as similarity search, information retrieval and question answering;
- Using NLP techniques in conjunction with computer vision to solve document understanding problems;
- I took the initiative to identify and solve problems in the development cycle of AI-related projects, leading to the development of a framework using dvc, Mlflow, git and docker for transversal use in all data science projects;
- Contact with clients and idealisation of AI projects with value, thinking of new ways to solve problems, resulting in new projects for the company;
- Contact with multi-faceted teams for the deployment and monitoring stage of AI projects.

Data Scientist Researcher *May 2020 – Mar 2022*
Centro Algoritmi

- Deep learning research in an autonomous driving project, for the automatic detection of audio anomalies in a self-driving taxi environment;
- Idealisation of a pipeline for mixing and synthesising synthetic data (audio), resulting in a hybrid dataset that allows for infinite combinations of anomalous and normal sounds, enabling the evaluation of algorithms for their detection;
- Development of generative adversarial networks (GANs) to automatically generate anomalous sounds (such as coughing and breaking glass) for later use in the tool mentioned above;
- Development of algorithms capable of detecting anomalies in sound clips in an unsupervised manner (various types of auto encoders, isolation forest and clustering methods);
- Opportunity to use the algorithms created on the basis of the synthetically generated dataset in a real context, in a live presentation;
- Use of open source technologies for versioning data and code, as well as managing pipelines (Mlfow, dvc and Dagshub).

Publications

Coelho, G., Matos, L. M., Pereira, P. J., Ferreira, A., Pilastri, A., & Cortez, P. (2022). Deep autoencoders for acoustic anomaly detection: Experiments with working machine and in-vehicle audio. *Neural Computing and Applications*, *34*(22), 19485–19499. https://doi.org/10.1007/s00521-022-07375-2

Coelho, G., Pereira, P., Matos, L., Ribeiro, A., Nunes, E. C., Ferreira, A., Cortez, P., & Pilastri, A. (2021). Deep dense and convolutional autoencoders for machine acoustic anomaly detection. In I. Maglogiannis, J. Macintyre, & L. Iliadis (Eds.), *Artificial intelligence applications and innovations* (pp. 337–348). Springer International Publishing. https://doi.org/10.1007/978-3-030-79150-6_27

Pereira, P. J., Coelho, G., Ribeiro, A., Matos, L. M., Nunes, E. C., Ferreira, A., Pilastri, A., & Cortez, P. (2021). Using deep autoencoders for in-vehicle audio anomaly detection [Knowledge-Based and Intelligent Information Engineering Systems: Proceedings of the 25th International Conference KES2021]. *Procedia Computer Science*, *192*, 298–307. https://doi.org/https://doi.org/10.1016/j.procs.2021.08.031

Skills

Languages: Portuguese (Native); English (Fluent)

Core Expertise: Machine Learning; Deep Learning; Natural Language Processing (NLP); Generative AI; Fairness & Privacy in AI; Responsible AI

Programming: Python; Pytorch; SQL; Bash

Frameworks & Libraries: PyTorch; TensorFlow; Scikit-learn; Transformers; DVC; MLflow; FastAPI

Tools & Platforms: Docker; Git; Azure (DevOps, ML); VectorDBs (e.g., FAISS, Weaviate); Jupyter; VS Code