The Nanoscience Cooperative Research Center, CIC nanoGUNE, located in Donostia / San Sebastian, Basque Country (Spain), is currently looking for a

MASTER STUDENT

to work on

Data Augmentation Techniques for Spectroscopic Data Using Physics-Aware and Generative Learning Approaches

NanoGUNE is a research center devoted to conducting world-class nanoscience research for a competitive growth of the Basque Country. NanoGUNE is a member of the Basque Research and Technology Alliance (BRTA) and is recognized by the Spanish Research Agency as a *María de Maeztu* Unit of Excellence.

The **position** is offered in the **Nanoengineering** group, led by Andreas Seifert. More information can be found at https://www.nanogune.eu/en/research/groups/nanoengineering

Project Description

Raman spectroscopy is widely used for chemical and material characterization across fields such as biomedical diagnostics, environmental monitoring, and material science. Machine learning methods are increasingly applied to automate and better understand the Raman spectra; however, the development of prediction models is often limited by the scarcity of available samples, especially in medical diagnostics, and variabilities introduced by noise, baseline drifts, and instrument-dependent distortions.

The focus of the proposed master's thesis is on the development and evaluation of advanced data augmentation strategies for Raman spectral data, but also data from other spectroscopic techniques, to improve the robustness and generalization capability of machine learning models. The **project** will explore both physics-inspired augmentations (e.g., peak shifting, intensity scaling, baseline variations, and noise modeling) and modern generative learning approaches for synthesizing realistic spectral data. Emphasis will be placed on novel or emerging techniques such as diffusion-based generative models, self-supervised augmentation learning, and physics-constrained generative frameworks that preserve chemically meaningful spectral features. The student will design and implement augmentation pipelines and evaluate their effectiveness in improving downstream tasks such as spectral classification, regression, or anomaly detection.

The project combines spectroscopy, machine learning, and signal processing, and aims to contribute to the development of robust data augmentation techniques for spectral datasets, which remain relatively underexplored compared to image and text domains.

Candidates should **apply** by following the instructions of the general call and by completing the form below and attaching the following documents:
   a. A complete CV
   b. Academic record and cover letter grouped in a single PDF file

The **deadline** for applications is **06/04/2026**.

(iii) We encourage you to subscribe to our [HR mailing list](#) to receive information related to nanoGUNE's open positions and open calls for different training and talent attraction programs.