Subject: Internship Project Proposal

Code Name	PTA_48652179_21/03/2025 20.41.40
Date	21/03/2025 20.41.40
Target	Bachelor's Degree - Data Analysis

Host Institution

Host Institution	Istituto Nazionale di Astrofisica (INAF)
Protocol	48652179
Protocol Date	10/01/2024
Country	Italy
City	Catania
Address	Via S. Sofia, 78, 95123
Website	https://www.oact.inaf.it/
Employees Number	323
Contant Person	Farida Farsian
Phone Number	+39 333 1461314
Email	farida.farsian@inaf.it

University of Messina, Italy
Department of Mathematical and Computer Sciences,
Physical Sciences and Earth Sciences

Project Supervisor

Name and Surname	Farida Farsian
Phone Number	+393331461314
Email	farida.farsian@inaf.it

Internship Project Details

Title	Analyzing Cosmic Structures with Graph Neural Networks

Detailed Description:
Context:
The large-scale structure (LSS) of the Universe forms a cosmic web of clusters, filaments, sheets, and voids. Understanding and classifying these structures is essential in cosmology. With the rise of machine learning in astrophysics, Graph Neural Networks (GNNs) offer a powerful tool to analyze spatial relationships directly from cosmological simulation data.

This project focuses on using the CAMELS simulation dataset, which provides dark matter halo catalogs with positions, masses, and other physical properties. The student will build graphs from halo catalogs, where halos are nodes connected based on spatial proximity, and apply a GNN model to classify halos into different cosmic environments (e.g., void, filament, sheet, cluster) based on local density or structural features.

Measurable Objectives:

- Successfully download and preprocess halo catalogs from the CAMELS dataset.
- Build a graph structure from halo positions (e.g., using k-Nearest Neighbors).
- Estimate the local density of halos and assign environment labels.
-Implement a GNN using PyTorch Geometric to predict the environment class of each halo.
- Train the model and evaluate its accuracy.
- Visualize the classification results in 3D space.

Tools & Libraries:

Dataset: CAMELS simulations (publicly available halo catalogs)
Programming Language: Python
Libraries:
PyTorch / PyTorch Geometric (GNN implementation)
NumPy / Pandas / scikit-learn (data processing)
Matplotlib / Plotly (visualization)
Optional: Jupyter Notebook for documentation and reporting

Topics	Application of Graph Neural Networks for Cosmic Environment Classification using CAMELS Simulations
Reimbursement of Expenses (YES/NO)	NO
Refund Amount	
Availability for Travel (YES/NO)	NO
Kind of employment	Part time
Duration in months (max 12)	9
Duration in hours	150
Internship Date Start	01/04/2025
Internship Date End	31/12/2025
Number of Open Position(s)	1

Internship Skills

Required Skills: Python programming (intermediate level)

Experience with:

NumPy / Pandas for data manipulation

scikit-learn for basic machine learning utilities

Familiarity with Jupyter Notebooks for coding and reporting

Desirable but not required:

Basic understanding of cosmology / large-scale structure

Other Skills	Self-learning ability: Willingness to explore unfamiliar tools like PyTorch Geometric and astrophysical datasets Problem-solving skills: Ability to debug code and address challenges during data processing and model training