Internship Announcement: Software Quality and LLMs-based Support Tools for Code in a Machine Learning Framework for Particle Physics at the CERN LHC

ACORN [1] is a cutting-edge, open-source Python framework dedicated to Geometric Deep Learning for Particle Physics detectors. It is being actively developed by the GNN4ITK [2-3] project team within the ATLAS [4] collaboration, part of the LHC experiment at CERN [5]. ACORN supports R&D for the design of a new charged particle tracking algorithm based on Graph Neural Networks (GNN). In this context of complex ML international R&D collaboration, sharing and evaluating models, as well as ensuring reproducibility of the results, are particularly important and rely on a high level of software quality.

We are offering an internship focused on significantly improving the software quality of the ACORN framework, particularly by enhancing its packaging, continuous integration and deployment processes, its tests and documentation. In addition to applying best practices, we plan to explore the use of innovative LLM-based tools to support code design, documentation generation, and other software quality engineering aspects.

The *Laboratoire des 2 Infinis – Toulouse* (L2IT) [6] is a young laboratory created in 2020 in Toulouse to conduct research in fundamental physics with new numerical and theoretical approaches to data analysis. The internship will be conducted within the *Computing, Algorithms, and Data* team at L2IT, under the direct supervision of a computer science research engineer.

You are a student in a Computer Science Master or in an Engineering School and you want to apply and develop your knowledge in Quality Software Engineering including using modern AI tools, this offer might be interesting for you! You will have the opportunity to enhance your skills while contributing to our R&D efforts in a highly stimulating and dynamic research environment as part of a major international scientific collaboration.

Example tasks for the Internship:

Software Quality Engineering:

- Designing and packaging Python code
- Developing unit and functional tests
- Managing repositories on GitLab and creating efficient GitLab workflows
- Implementing Continuous Integration (CI) pipelines
- Building Docker images with GPU support

Exploring LLM-Based Tools for Code and Documentation Generation:
- Conducting a state-of-the-art analysis of available tools
- Utilizing these tools to generate and design code, documentation, test cases, and CI configuration files

Required Skills:

- Proficiency in Python development
- Notion in software quality best practices, version control (Git), GitLab, and Docker
- English language proficiency at a B2 level or higher
- Experience with Numpy, Pandas, and PyTorch is a plus

Contacts:

Sylvain Caillou, sylvain.caillou@l2it.in2p3.fr
Maxime Pigou, maxime.pigou@l2it.in2p3.fr

References:

[1] GNN4ITk Team, A Charged-particle geOmetric Reconstruction Network (ACORN), 2023, url: https://gitlab.cern.ch/gnn4itkteam/acorn

[2] Sylvain Caillou et al. ATLAS ITk Track Reconstruction with a GNN-based pipeline. Tech. rep. 2022. URL: https://cds.cern.ch/record/2815578.

[3] Caillou, Sylvain et al. "Physics Performance of the ATLAS GNN4ITk Track Reconstruction Chain"

[4] ATLAS experiment, url: https://atlas.cern

[5] CERN, url: https://home.cern/

[6] L2IT, url: https://www.l2it.in2p3.fr