Internship SI4 Summer 25

Interactive Visualization in Machine Learning

Context

Machine Learning (ML) workflows are becoming increasingly complex, making their structure and reuse difficult and cumbersome. Many ML workflows are hidden within code, making it challenging to understand, validate, and reuse them effectively. Ensuring the correctness and maintainability of these workflows is crucial for improving best practices and enabling efficient collaboration between developers and data scientists.

The **PROFIL** research project, funded by the French National Research Agency, aims to extract, visualize, and analyze the **profiles** of ML notebooks. These profiles provide an overview of the key stages in an ML workflow, such as:

- **Data preparation**
- **Model training**
- **Results visualization**

However, manually identifying these steps can be complex. This internship will focus on developing and enhancing an **interactive visualization tool**, allowing users to explore, build, and understand these notebook profiles according to their interests.

In addition, this internship will also be part of a second ANR-funded project, **FATEsMLOps**, which aims to verify and guide developers in addressing Fairness, Accountability, and Transparency properties in ML applications. The hosting team is **SPARKS** at **I3S in Sophia Antipolis**, integrating both the software engineering team MUSCAT and part of the **MAASAI team at INRIA**. The intern will also share the national and international settings of the current research projects, through collaborations with team members from **Lille, Toulouse, and McMaster University in Canada**.

Internship Objectives

The intern will actively contribute to the development of the first components of this visualization tool, interacting with data scientists and potentially non-specialized developers.

Main Responsibilities

1. **Understanding and Enhancing the Existing User Interface** (developed in React) based on user feedback and project progress. The goal is to create a version of the tool that supports scientific exploration of notebooks for both novices and experts.
2. **Integration with Core Components**, including:
    - A profile repository (subject to further evolution)

    - A workflow pattern repository
    - Profile extraction components using a parser under development and a dedicated LLM (Large Language Model).
3. **Deployment and Online Availability** of the notebook evaluation environment, with tracking of user interactions and usage. A **DevOps/MLOps approach** will be followed.

Technologies and Environment

The intern will work with the following technologies:

- **React** for the user interface development.
- **Docker** for deploying various components.
- **Python** for core components, data transformations, and interactions.
- A **Git** code repository platform with associated automation.

Desired Profile

We are looking for a student with a strong **software engineering background** who is passionate about UI development and curious about ML-related challenges. A good understanding of Web development principles, a collaborative mindset, and attentiveness to user needs are expected.

Why Join This Internship?

- **An innovative project** at the heart of current ML industrialization challenges
- **Interdisciplinary experience**, working with researchers in software engineering and machine learning
- **A chance to make an impact** by contributing to a tool that could be widely adopted