ASPETTANDO LA
NOTTE EUROPEA DEI RICERCATORI

Reimagining Industrial Automation Programming with AI: The ST Copilot Project

Adnan Riaz

Department of Computer Science and Engineering - PhD in Computer Science and Engineering - 38th ciclo - Moreno Marzolla

Background

- **Industrial Automation and the Role of PLCs** – PLCs are the backbone of industrial automation, enabling reliable control of machines and processes.
- **The Significance of Structured Text** – Structured Text (ST) is a high-level IEC 61131-3 language crucial for writing complex logic in industrial automation systems.
- **Rise of AI and Large Language Models in Code Generation** – LLMs are transforming software engineering by automating code generation and assisting developers across domains.
- **Challenges in Applying LLMs to Structured Text Code** – Lack of public datasets for ST code, absence of ST benchmarks and evaluation protocols, limited tool support for automated validation.

Project Goals

The main project's goals/objectives are described below.

- Creation of Structured Text Dataset.
- Evaluate existing code LLMs and fine-tuned models.
- Methods for low computational fine-tuning and domain adaptation.
- Address licensing and compliance challenges in deploying LLMs for industrial automation.
- Explore the practical applications and implications of LLM-assisted code generation in industrial automation.

Experimental Approach

- A. **Model Architecture**→ pre-trained phi-3-mini-128k-instruct.
- B. **Parameter Efficient Fine Tuning (PEFT)**→ "freezes" most layers and only trains the last few layers, requiring lower computational resources and less time.
- C. **Low Ranked Adaptations (LoRA)**→ PEFT fine-tuning technique. LoRA Alpha =16, r =32, Bias= none, Dropout =0.05, target modules = qkv proj, o proj, gate up proj, down proj.
- D. **Dataset**→ ST dataset, 200K problem.
- E. **Training** → Amazon Web Service (AWS) cloud infrastructure (g5.xlarge EC2 instance with 24GB GPU Memory).

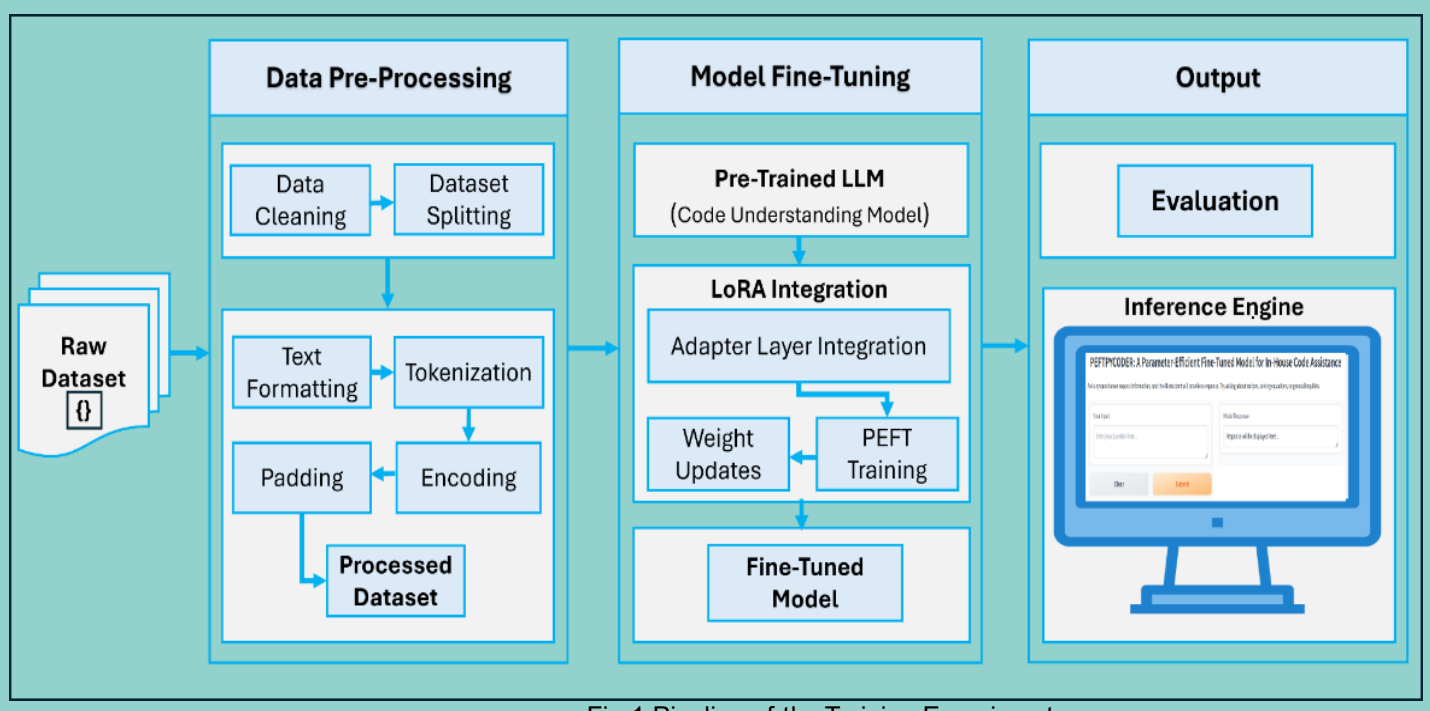

Fig.1 Pipeline of the Training Experiment

Expected Outcomes

- **Robust ST Dataset** – Creation of a large, diverse, and validated dataset (200K Examples) of Structured Text (ST) programs to fill the current gap in public resources.
- **Automated Pipeline for ST Code Generation** – Development of a scalable pipeline that leverages LLMs to generate, parse, and validate ST code with iterative feedback.
- **Integration of Semantic Validation** – Implementation of automated testing frameworks (e.g., TcUnit) to ensure generated code is not only syntactically correct but also semantically accurate.
- **Benchmarking Framework** – Establishment of evaluation protocols and metrics tailored for ST code generation, enabling systematic comparison of models.
- **Demonstration of Low-Resource Feasibility** – Evidence that effective ST code generation can be achieved with lightweight or fine-tuned models, making deployment practical for industry use.
- **Enhanced Understanding of LLM Capabilities for Automation** – Insights into the strengths, limitations, and improvement areas of LLMs when applied to industrial automation programming.

SUSTAINABLE DEVELOPMENT GOALS