Internship Project

Evaluating Large Language Models on Time Series Analysis Tasks

Supervisors	*Paul Boniol* and *Roberto Stanzione*
Contacts	*paul.boniol@inria.fr* *roberto.stanzione@inria.fr*
Host team	VALDA team
Location	École Normale Supérieure, 45 rue d'Ulm, Paris
Start date	March 2027 (at the latest)
Duration	6 months

Context and Motivation

Large language models (LLMs) are increasingly being used to analyze time series, yet their ability to capture and reason about temporal data, compared with human analysts and state-of-the-art methods, remains largely unexplored. It is also unclear how their performance is affected by the way time series are presented (e.g., through visual or textual representations) and by potential additional information provided alongside them.

Objectives

This project aims to systematically evaluate the capabilities of several frontier LLMs, including reasoning-enabled models, across three fundamental time series mining tasks: anomaly detection, motif discovery, and segmentation. The performance of LLMs will be compared against both state-of-the-art algorithms and human analysts, across diverse datasets and task-specific questions.

Methodology and Expected Outcomes

To this end, this project aims to develop a rigorous experimental framework and a controlled evaluation protocol to enable fair and systematic comparisons across models, algorithms, and human participants, ultimately providing a comprehensive assessment of the strengths and limitations of current LLMs in time series analysis.

Practical Information

The internship will take place within the VALDA team, located at the École Normale Supérieure, 45 rue d'Ulm, Paris. The project will start in March 2027 at the latest and will last 6 months.