3iA Côte d'Azur
Interdisciplinary Institute for Artificial Intelligence

Ph.D. research topic

- Title of the proposed topic:
  Online change-points and events detection for dynamic structured data
- Research axis of the 3IA: AI FOR SMART AND SECURE TERRITORIES
- **Supervisor (name, affiliation, email):** **Cédric RICHARD, Lagrange, UCA**
  cedric.richard@unice.fr
- The laboratory and/or research group: Lagrange Lab, UCA

Apply by sending an email directly to the supervisor.
The application will include:

- Letter of recommendation of the supervisor indicated above
- Curriculum vitæ
- Motivation letter
- Academic transcripts of a master's degree(s) or equivalent
- At least, one letter of recommendation
- Internship report, if possible

- Description of the topic:

Change-point detection (CPD) is a fundamental problem in statistics and machine learning, focused on identifying abrupt shifts in the properties of data over time. These shifts, known as change-points, indicate transitions in the underlying distribution or dynamics of a system, which may result from external events or internal structural changes. The objective of CPD is to pinpoint when these changes occur and, in some cases, to understand the nature of the shifts. CPD has a wide range of applications across domains that require online insights and adaptive decision-making, such as medical monitoring, real-time trading, and network security. A growing number of these applications are generating structured, high-dimensional data with non-trivial and intricate geometric properties. These data often display complex relationships and dependencies that go beyond Euclidean spaces, necessitating sophisticated techniques for analysis and interpretation. Prominent examples include time sequences on groups and manifolds, time sequences of graphs, graph signals, etc.

A major challenge with dynamic structured data is finding representations that can effectively handle their underlying geometry, which is often defined by application-specific pseudo-distances. A common approach is to embed such data into conventional geometric spaces, like Euclidean spaces, even when the data may be more naturally represented in non-Euclidean domains. This mismatch in representation complicates both learning and inference processes. Another challenge is that dynamic structured data are generated by a variety of sensors and infrastructures that continuously produce, disseminate, and store information. However, this data deluge already surpasses our capacity for analysis and decision-making, necessitating online actions rather than offline processing due to its

time-sensitive nature. The monitoring of telecommunications and energy production and distribution networks are characteristic examples of such time-critical applications.

The project aims to propose unsupervised online CPD algorithms for dynamic structured data, with a particular focus on time sequences on groups and manifolds. Non-parametric frameworks will be specifically considered, as they make fewer, if any, assumptions about the data's underlying distribution and are better suited for detecting a wider variety of changes. The CPD algorithms will be designed to be computationally efficient to ensure scalability. Finally, an open-world recognition setting will be developed to classify changes on-the-fly into either previously seen classes or unknown classes. Applications to smart cities monitoring are considered.