

AISSAI Heterogeneous Data and Large Representation Models in Science Workshop

Toulouse (France), September 30 - October 3 2024

FIRST ANNOUNCEMENT

The third workshop of the [AISSAI trimester on "Artificial Intelligence for the two infinities"](#) will be held in Toulouse (France) from September 30th to October 3rd, 2024. This workshop will be devoted to *Heterogeneous Data and Large Representation Models in Science*.

This event is sponsored by the [CNRS AISSAI center](#) and hosted by the [Laboratoire des 2 Infinis – Toulouse \(L2IT\)](#).

Please find complete information on the scientific program, call for abstracts and registration on the [workshop's Indico website](#). Registration is free (but mandatory) and includes lunches and social events.

Scientific program

In recent years, we have witnessed remarkable transformations in the AI/ML landscape. Particularly in computer vision and natural language processing, there is a notable emergence of Large Representation Models (LRMs) trained on extensive datasets, often referred to as foundation models. These LRMs possess the capability to encode information at a high level of abstraction, enabling the training of models on multimodal data such as text, image, sound, video, and more. This improvement in how models represent objects augments significantly their ability to understand and make sense of the world.

In the realm of science, we expect a similar revolution, triggered by the integration of heterogeneous multimodal scientific data from various sensor systems or sources into LRMs. Scientific data are often heterogeneous and multimodal in nature, originating from various sensors in embedded systems (robotics, aerospace), from different detector subsystems or different instruments in fundamental physics, or from different signal sources in a scientific experiment in general. The models can combine representations from neural networks with symbolic representations integrating a priori knowledge of the scientific domain.

The aim of this workshop is to bring together scientists from different fields (just to list a few: computer sciences, cosmology, human sciences, mathematics, physics, robotics, statistics, etc.) and with different profiles (experimentalists, theorists, developers) to discuss these topics at the forefront of AI/ML research, fostering collaboration and innovation in this rapidly evolving field.

The program is planned to be a mix of high-profile guest speaker presentations and contributed talks.

This workshop will delve into a range of topics, which include but are not limited to:

- Constructing machine learning models capable of learning from diverse data types.
- Managing multimodal data from varied sources, or heterogeneous data from scientific instruments that combine multiple detector technologies, for ML applications.
- Investigating contrastive embeddings tailored for heterogeneous and multi-modal scientific data alongside shared embedding representations.
- Exploring the integration of neuro-symbolic AI and multi-level representations.
- Mathematical modeling of combined representation.

- Exploring explainability and interpretability of Large Representation Models in the scientific context.
- Embracing frugality and size management in Large Representation Models.
- Possibly on a longer timescale, exploring numerical encodings for large language representations in scientific contexts.

Website (Indico)

<https://indico.in2p3.fr/event/33412/>

Important dates

These dates may change.

Registration opening:	open now
Abstract submission opening:	open now
Abstract submission dead-line:	August 15 th
Abstract acceptance dead-line:	August 31 th
Program release:	September 7 th
Registration closing	September 15 th
Workshop start	September 30 th

Scientific advisory committee

- Sylvain Caillou (L2IT, IN2P3, CNRS/UT3) - chair
- Alexandre Boucaud (APC, IN2P3, CNRS)
- Tobias Golling (Université de Genève)
- François Lanusse (Polymathic AI)
- Daniel Murnane (Copenhagen University)
- Thomas Oberlin (ISAE-SUPAERO, ANITI, Université Paul Sabatier)
- Jan Stark (L2IT, IN2P3, CNRS/UT3)
- Gordon Watts (Washington University)

Local organisation committee

- Catherine Biscarat (L2IT, IN2P3, CNRS/UT3) - chair
- Sylvain Caillou (L2IT, IN2P3, CNRS/UT3)
- Jocelyne Gauthier (L2IT, IN2P3, CNRS/UT3)
- Jan Stark (L2IT, IN2P3, CNRS/UT3)
- Jeanette Thibaut (L2IT, IN2P3, CNRS/UT3)