https://indico.in2p3.fr/e/AISSAI-TLS

AI SSAi
AI for science, science for AI

Heterogenous Data and Large Representation Models in Science

30 September 2024 to 3 October 2024, Toulouse (L2IT)
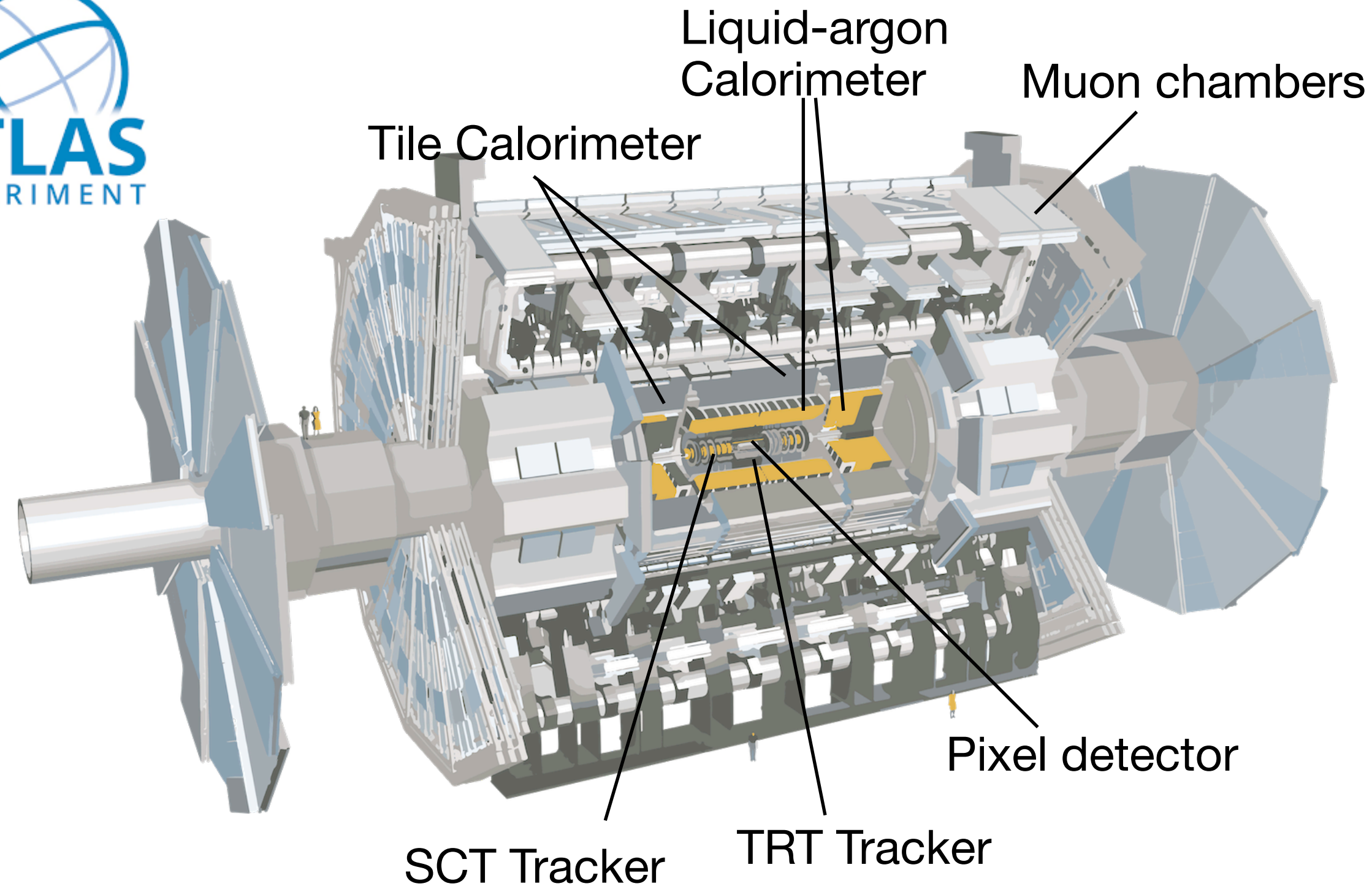
# The ATLAS detector and AI

ATLAS, one of the major detectors at CERN's Large Hadron Collider (LHC)

ATLAS has the dimensions of a cylinder, 46m long, 25m in diameter

Currently a lot of Machine Learning (ML) is used in each detector subsystems (InnerTracker, calorimeter, …)

The ATLAS detector and AI

Liquid-argon Calorimeter

Muon chambers

Tile Calorimeter

Pixel detector

SCT Tracker

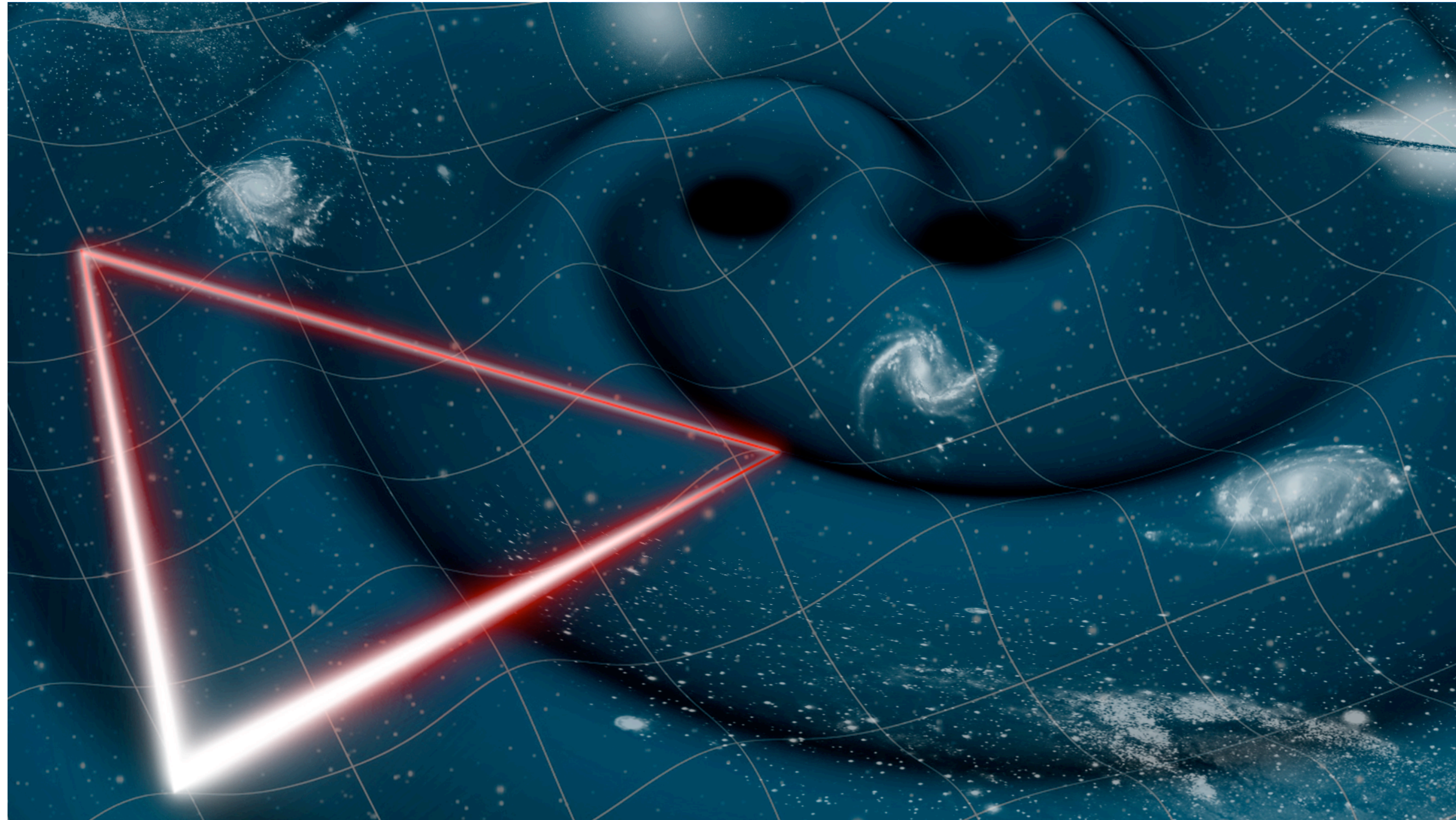TRT Tracker
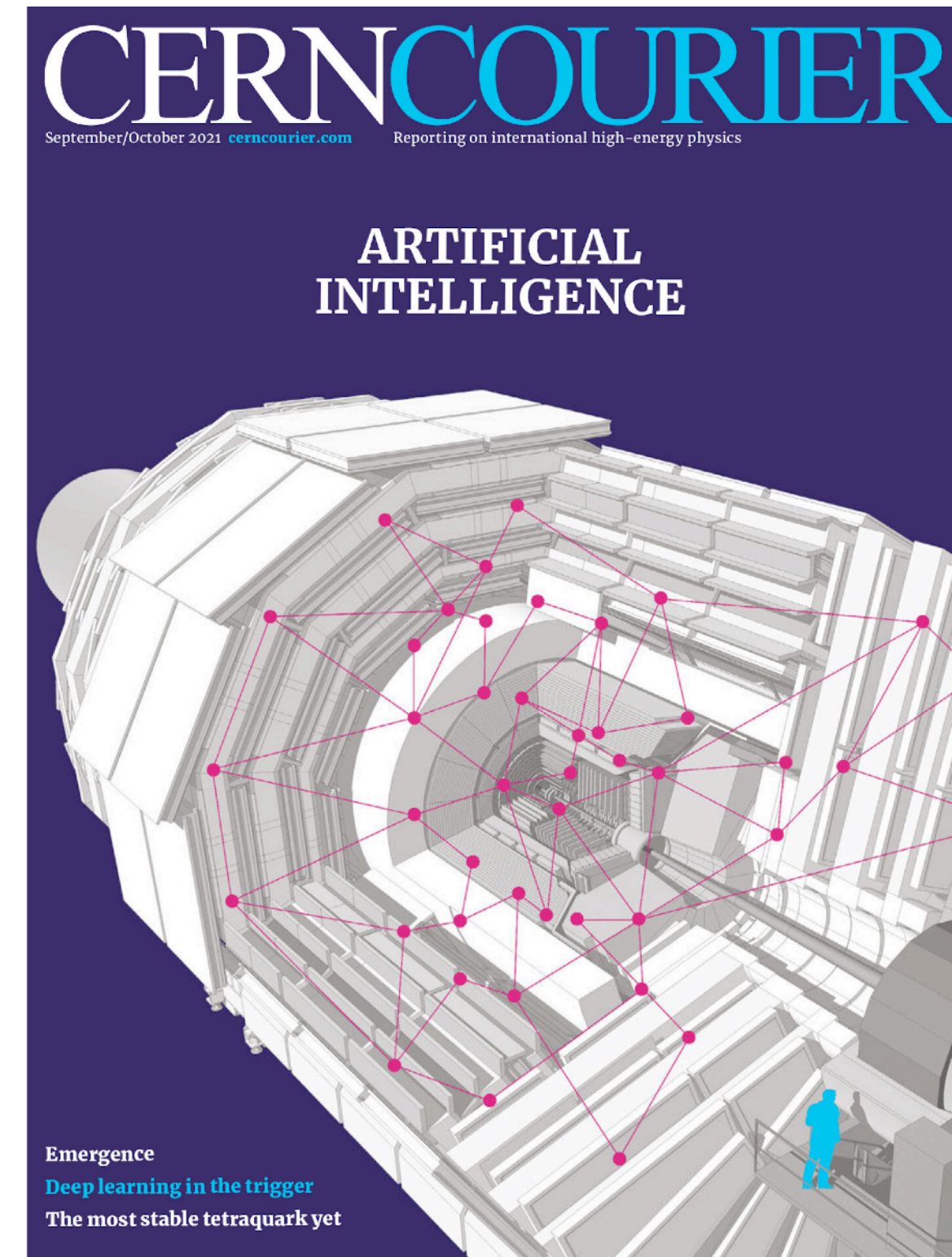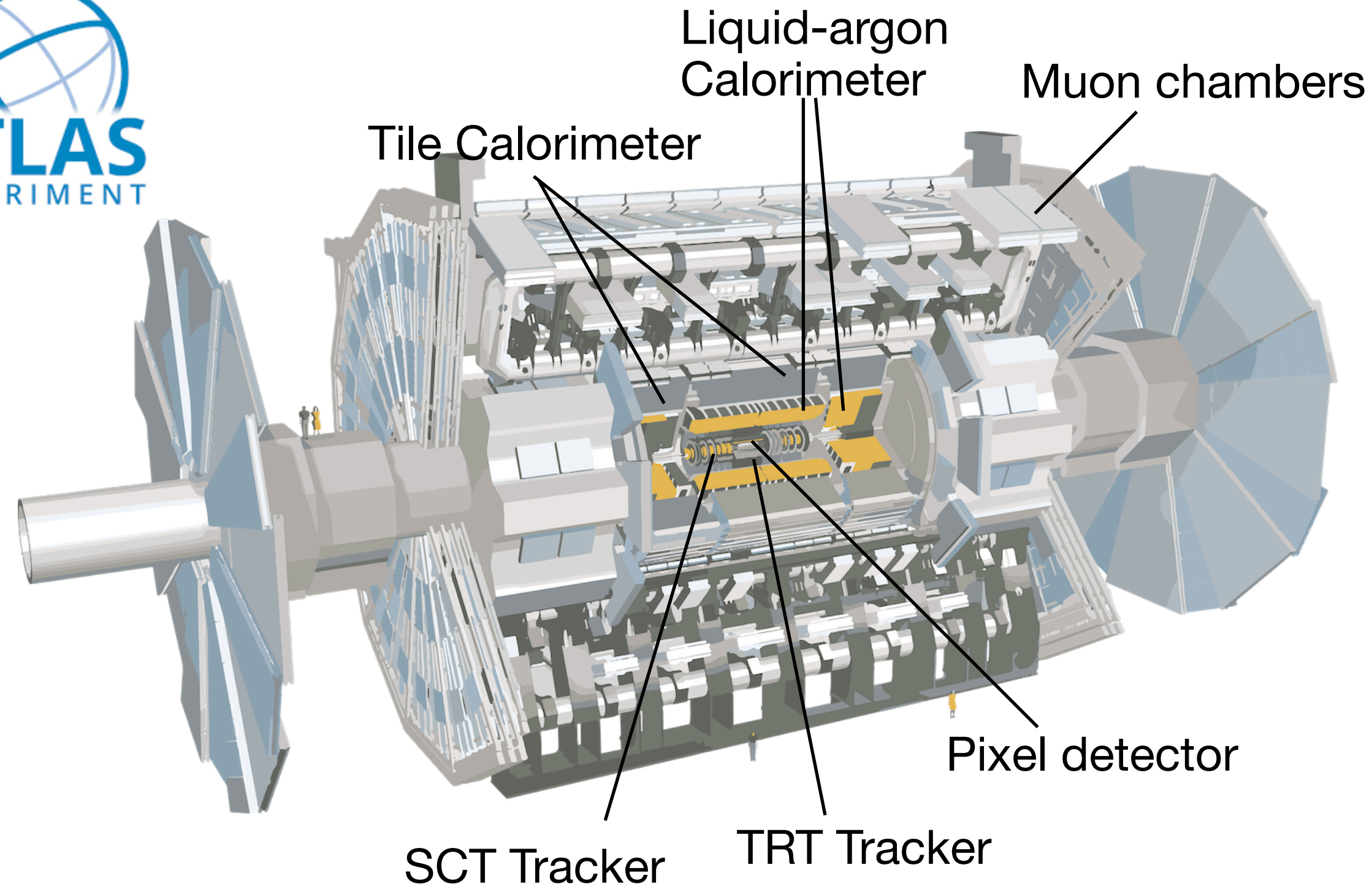
How to learn from Heterogenous and Multimodal data from different detector subsystems ?

How can AI models can combine Heterogenous and Multimodal data to get a better representation of the physic object ?

AI for science, science for AI

Laser Interferometer Space Antenna (LISA)

- Future Gravitational Waves space detector

- 3 spacecraft arranged in an equilateral triangle with each side 2.5M km long

- Signal will be thousand of Gravitational Waves coming from everywhere from hetereogenous sources (galactic binaries, mergers of massive black holes, …)

Separating overlapping Gravitational Waves signals is an *extremely* hard problem (global fit challenge)

Do AI can help with that ?

AI for science, science for AI

Laser Interferometer Space Antenna (LISA)

Penguins use distinct vocal signatures (combination of frequency, amplitude, and temporal patterns) to identify individual calls amid the colony's cacophony

Able to find mates and offspring even in noisy, crowded environments.
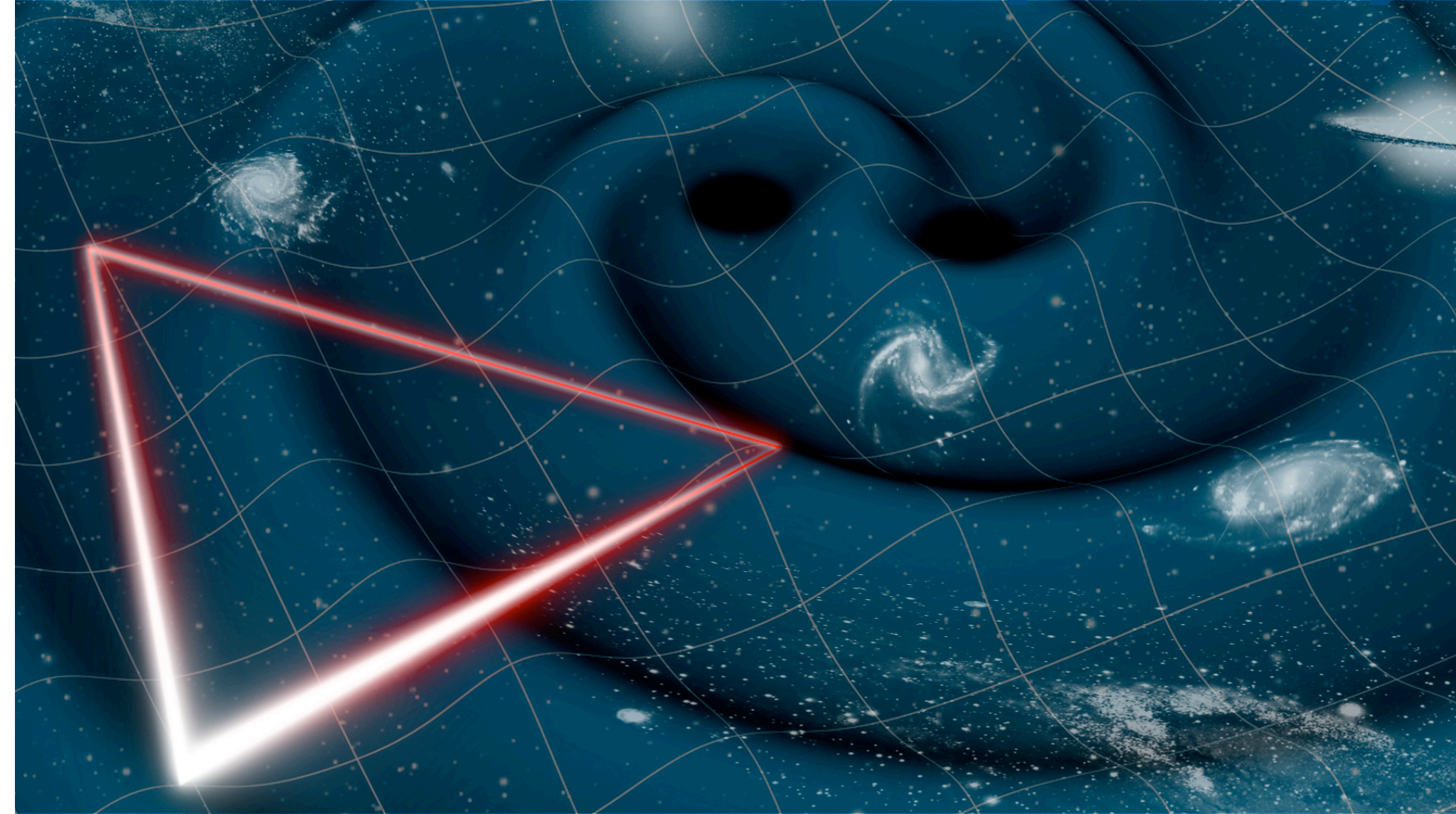
How does it compare to the global fit in LISA ?

If Natural Intelligence can do it, can Artificial Intelligence do it as well?

AI for science, science for AI

Artificial intelligence in a new era

‣ Rise of Large Language Models (LLMs)

‣ Multimodal Generative AI: Text + Image + Videos,…

‣ Retrieval Augmented Generation (RAG): Mix LLMs with knowledge databases

‣ Neurosymbolic AI: Integrate Neural Network-based methods with symbolic knowledge based approaches

‣ Rise of Representation / Geometric Learning (Graph Neural Networks): Generalisation of learning from structural patterns in data

• Likelihood-free Learning: Solving complex non-linear inverse problems with unknown and complex prior distribution and in a noisy and / or multiple heterogeneous sources environment

Tomorrow's AI foundation will be capable of learning a high level of abstraction in representation from heterogeneous and multimodal data of various types and nature.

How does this apply in science?

AI for science, science for AI

At the origin of the workshop

[AISSAI](#) (AI for Science, Science for AI) is a CNRS interdisciplinary center that promotes connexions between scientific domains working with AI.

The IN2P3 has been chosen to organise the third AISSAI semester on "Artificial Intelligence for the two infinites physic"

AI and the uncertainty challenge in fundamental physics

Anomaly Detection Workshop

Heterogenous Data and Large Representation Models

At the origin of the workshop

AISSAI (AI for Science, Science for AI) is a CNRS interdisciplinary center that promotes connexions between scientific domains working with AI.

The IN2P3 has been chosen to organise the third AISSAI semester on "Artificial Intelligence for the two infinites physic"

AI and the uncertainty challenge in fundamental physics

Anomaly Detection Workshop

Heterogenous Data and Large Representation Models

At the origin of the workshop

AISSAI (AI for Science, Science for AI) is a CNRS interdisciplinary center that promotes connexions between scientific domains working with AI.

The IN2P3 has been chosen to organise the third AISSAI semester on "Artificial Intelligence for the two infinites physic"

AI and the uncertainty challenge in fundamental physics

Anomaly Detection Workshop

Heterogenous Data and Large Representation Models

Bringing people together

NATURAL INTELLIGENCE

LINGUISTIC

PARTICLE PHYSICS

COGNITIVE SCIENCES

BIOLOGY

EARTH OBSERVATION

ACCELERATORS

MATHEMATICS

ANITI

NEUTRINO PHYSICS

ARTIFICIAL INTELLIGENCE

MEDICAL IMAGING

COSMOLOGY

GRAVITATIONAL WAVES

SATELLITES

ASTROPHYSICS

FLUID MECHANICS

SPACE

AERONAUTICS

AI for science, science for AI

Thematic sessions

- Foundation models in Science

  ‣ Monday afternoon

- Heterogenous Data and Multimodal Representation Learning

  ‣ Tuesday afternoon & Wednesday second half of morning and afternoon

- Non linear Inverse Problem - Likelihood-free Learning

  ‣ Tuesday morning (the specific case of LISA) & Wednesday morning

Thanks to the Scientific Advisory Comittee

François Lanusse
(Polymathic AI)

Alexandre Boucaud
(APC, IN2P3, CNRS)

Daniel Murnane
(Copenhagen University)

Sylvain Caillou
(L2IT, IN2P3, CNRS/UT3)

Thomas Oberlin
(ISAE-SUPAERO, ANITI,
Université de Toulouse)

Tobias Golling
(Université de Genève)

Gordon Watts
(Washington University)

Jan Stark
(L2IT, IN2P3, CNRS/UT3)

AI for science, science for AI

Thanks to the Scientific Advisory Comittee

François Lanusse
(Polymathic AI)

Alexandre Boucaud
(APC, IN2P3, CNRS)

Daniel Murnane
(Copenhagen University)

Sylvain Caillou
(L2IT, IN2P3, CNRS/UT3)

Thomas Oberlin
(ISAE-SUPAERO, ANITI,
Université de Toulouse)

Tobias Golling
(Université de Genève)

Gordon Watts
(Washington University)

Jan Stark
(L2IT, IN2P3, CNRS/UT3)

AI for science, science for AI

14

Heterogenous Data and Large Representation Models in Science WS

AI for science, science for AI

PRATICALITIES

Thanks to the "Village" to welcome us

WiFi Instructions

SSID: EventVillageByCA

Password: V1ll@ge0313

Scientific program

On indico

AISSAI - Heterogeneous Data and Large Representation Models in Science
from **Monday 30 September 2024 (12:30)** to **Thursday 3 October 2024 (14:25)**

Monday 30 September 2024

Time	Session
12:30	Registration & light lunch
14:00	Welcome from the L2IT director - Jan Stark (L2I Toulouse, CNRS/IN2P3, UT3)
14:10	Welcome from the organizers; practicalities - Sylvain Caillou (L2I Toulouse, CNRS/IN2P3, UT3)
14:25	Keynote Address: Foundation models for high energy physics
15:15	PolarBERT: a Foundation Model for Neutrino Telescope Data
15:50	Coffee break
16:20	Scientific Foundation Models for Computational Fluid Dynamics: threats and opportunities
16:45	Small thinks big: transfer learning in KM3NeT/ORCA for neutrino event reconstruction
17:25	Automatic estimation of the wind turbine noise with recurrent neural networks
17:45	Free time
19:00	Welcome cocktail

Tuesday 1 October 2024

Time	Session
09:30	Keynote Address: Gravitational waves coming at you from all directions - Jonathan Gair (AEI Potsdam)
10:15	Coffee break
10:45	Statistically principled learning for gravitational-wave inverse problems
11:10	Neural density estimation for Galactic Binaries in LISA data analysis - Natalia Korsakova (APC)
11:35	Beyond Gauss? A more accurate model for LISA astrophysical noise sources
12:10	Group picture
12:30	Lunch
14:00	Keynote Address: Deep learning and the global workspace theory
14:15	Semi-supervised multimodal representation learning through a global workspace
14:50	Learning how to design biomolecules using a neuro-symbolic architecture
15:30	A graph-structured distance for heterogeneous datasets with meta variables
16:05	Challenges of heterogeneous data for building Linguistic Theory - Anisia Popescu (LISN)
16:25	Coffee break
16:55	Allocated - round table

Wednesday 2 October 2024

Time	Session
09:00	Keynote Address: RELEO - Representation Learning for Earth Observation - Jordi INGLADA (CESBIO, Université de Toulouse, CNES/CNRS/INRAe/IRD/UT3)
09:50	Identifying a piecewise affine signal from its nonlinear observation - application to DNA rep...
10:25	Coffee break
10:55	Keynote Address: Multimodal Pretraining for Astrophysical Foundation Models - François Lanusse (CNRS, UMR AIM / Flatiron Institute)
11:45	Galaxy detection with deep learning in radio data - David Cornu (Observatoire de Paris I PSL)
12:25	Searching for Dark Matter at the LHC with GNN - Rafal MASELEK (LPSC (Grenoble))
12:35	Architecture for intelligent web service in the Web of things
12:40	Lunch
14:00	Explaining Jet Flavour Taggers with Integrated Gradients
14:25	Salt: Multimodal, Multitask Models for the ATLAS Experiment - Jackson Barr (UCL)
15:05	Graph Neural Networks for track reconstruction in the ATLAS ITk detector
15:40	Poster coffee break
16:10	Large-scale deep-learning for weather and climate prediction - Laure Raynaud (Météo-France)
16:50	Keynote Address: Medium Range Weather Forecasting with Machine Learning
17:35	Free time
19:30	Workshop diner

Thursday 3 October 2024

Time	Session
09:00	Space is available for your luguages.
09:30	Enhancing Ultrasound Localization Microscopy (ULM) with Spatio-Temporal Deep Learning
09:55	Preprocessing arbitrarily structured data for AI with Awkward Array
10:35	Leveraging AI in computational physics with NVIDIA Modulus and TorchFort
11:10	Coffee break
11:40	Optimizing PyTorch: Accelerating Training and Inference with Compilation, Custom Kernel...
12:10	Keynote Address: A causal perspective on reliable and interpretable representation learning - Michel Besserve (Max Planck Institute for Intelligent Systems)
12:55	Closing words / farewell
13:00	Lunch

AI SSA I
AI for science, science for AI

Talks

- All talks are taking place in the auditorium

- Keynote invited speaker talk: 35'+10'

- Middle talk: 25'+10'

- Short talk: 15'+5'

- The speaker will be warned before the end of their talk by timing signs

- Please respect your speaking time

Talks will be recorded, with the intent to put them on the web.

▸ Let us know if you would not like to be recorded.

Plenty of discussion time, do not hesitate to ask questions

Speaker Support

For those of you who have an account on the IN2P3 Indico server you can self-upload your slides through Indico:

- Logging into the event: https://indico.in2p3.fr/event/33412/overview

- Click on 'My Contributions' toward the bottom of the Menu sidebar on the right

  ➡This will only show up if you are logged in!

- Click on the talk.

  ➡Upload the talk from the 'Pencil' icon
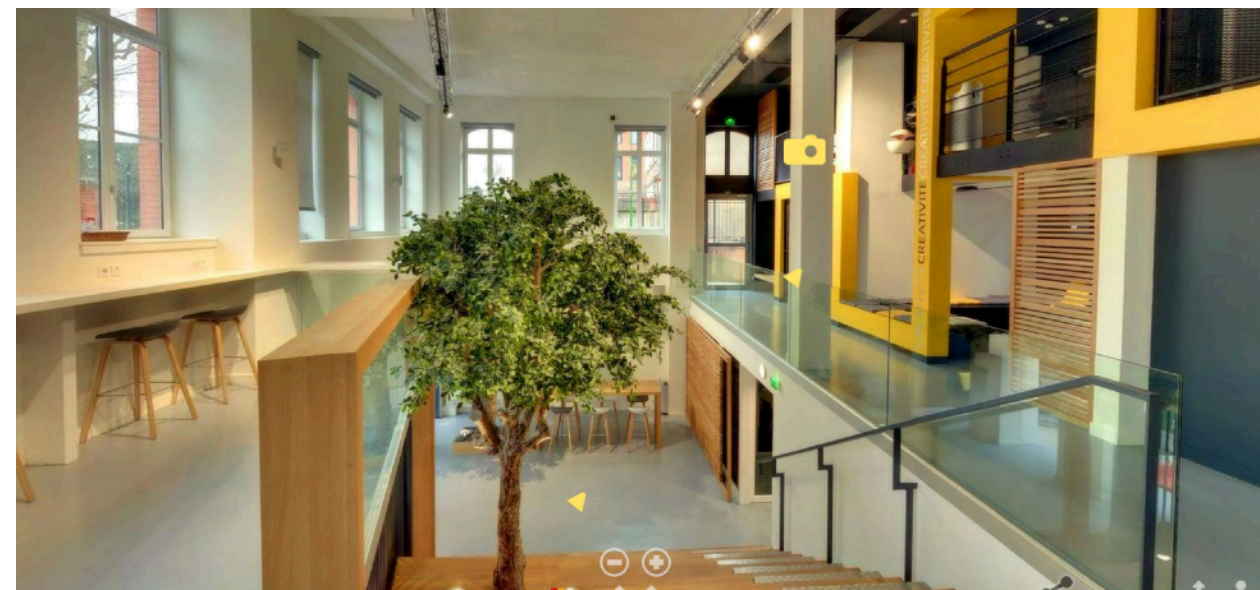
If you do NOT have an account, please send your slides to sos@l2it.in2p3.fr

- Please upload your slides at least 20min before the start of the sessions!

- PDF ONLY

Coffee and lunch breaks

Coffee (morning & afternoon) and Lunch breaks are taking place in « la place du village » (the door opposite to the auditorium entrance)

We will have lunch both on the ground floor and in the area down the stairs. Please spread out!

Social events

L2IT

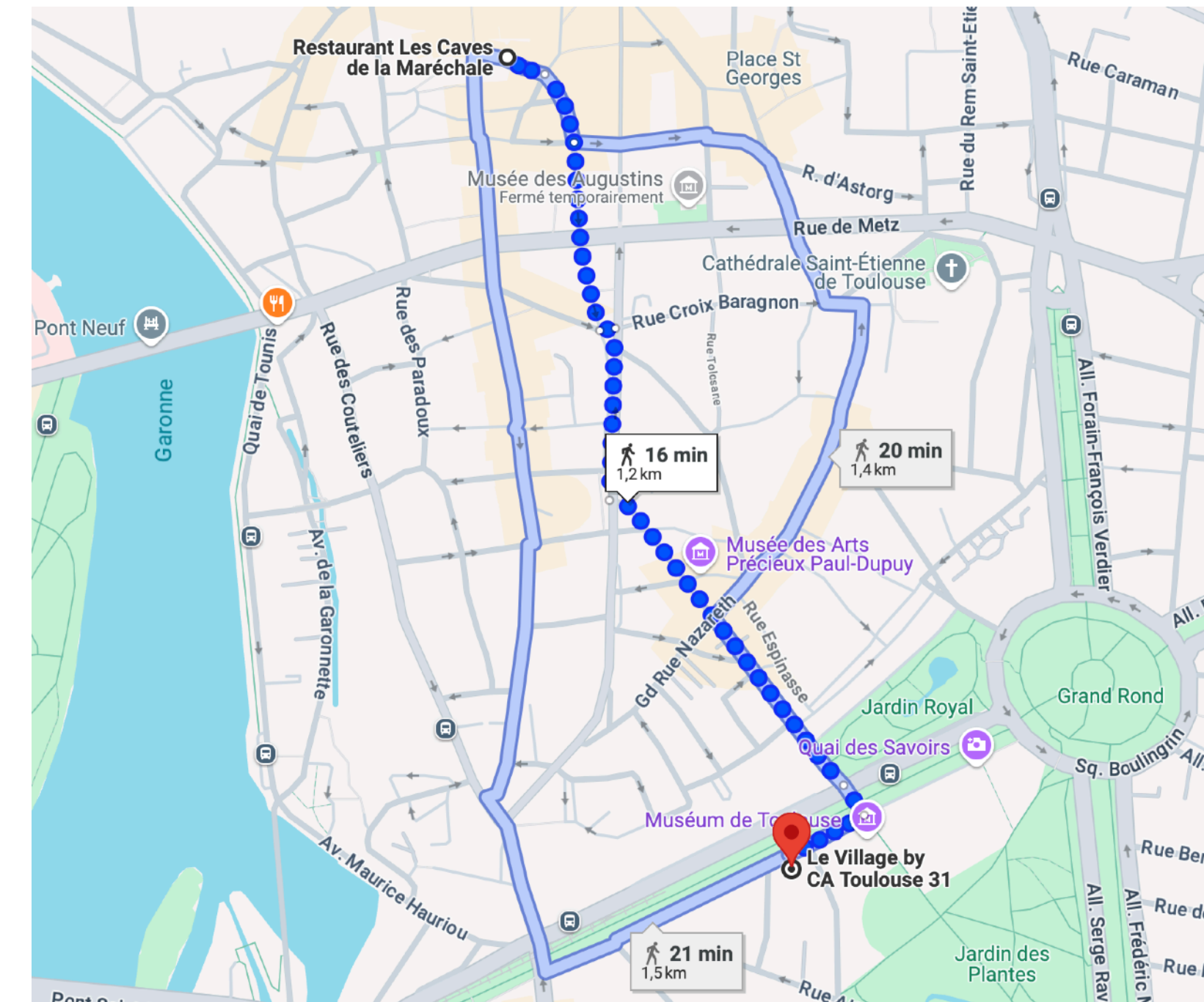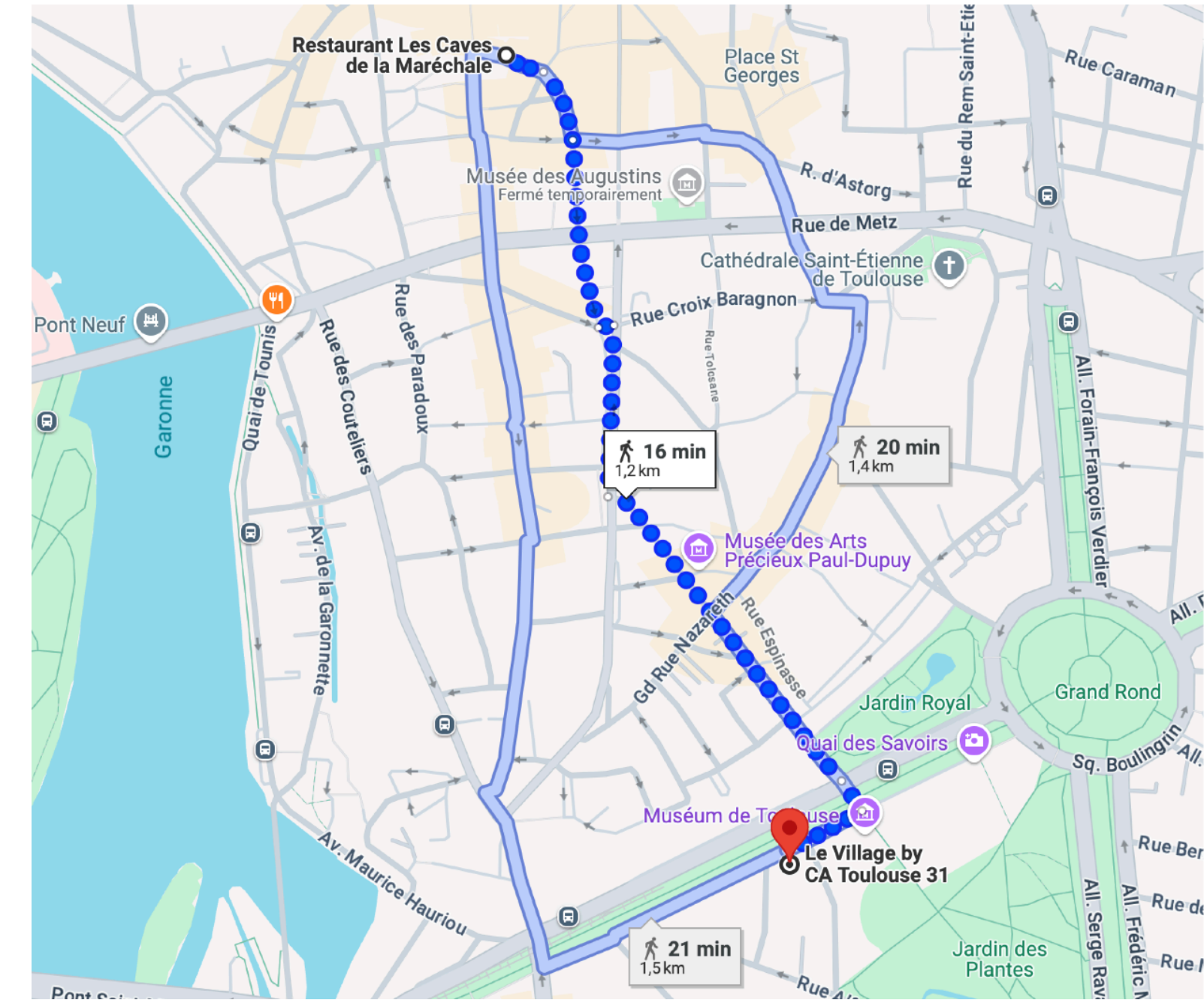

Bienvenue cocktail

Tonight @7pm

at The "Place du Village"

Workshop diner

Wednesday @7:30pm

At *Les Caves de la Maréchale*

3 rue Jules Chalande 31000 Toulouse

AISSAI
AI for science, science for AI

Social events

Bienvenue cocktail

Tonight @7pm

at The "Place du Village"

Workshop diner

Wednesday @19:30pm

At Les Caves de la Maréchale

3 rue Jules Chalande 31000 Toulouse

Have you already informed us of your diner menu choice ?
If not, please do so now

AI for science, science for AI

Social events

Bienvenue cocktail

Tonight @7pm

at The "Place du Village"

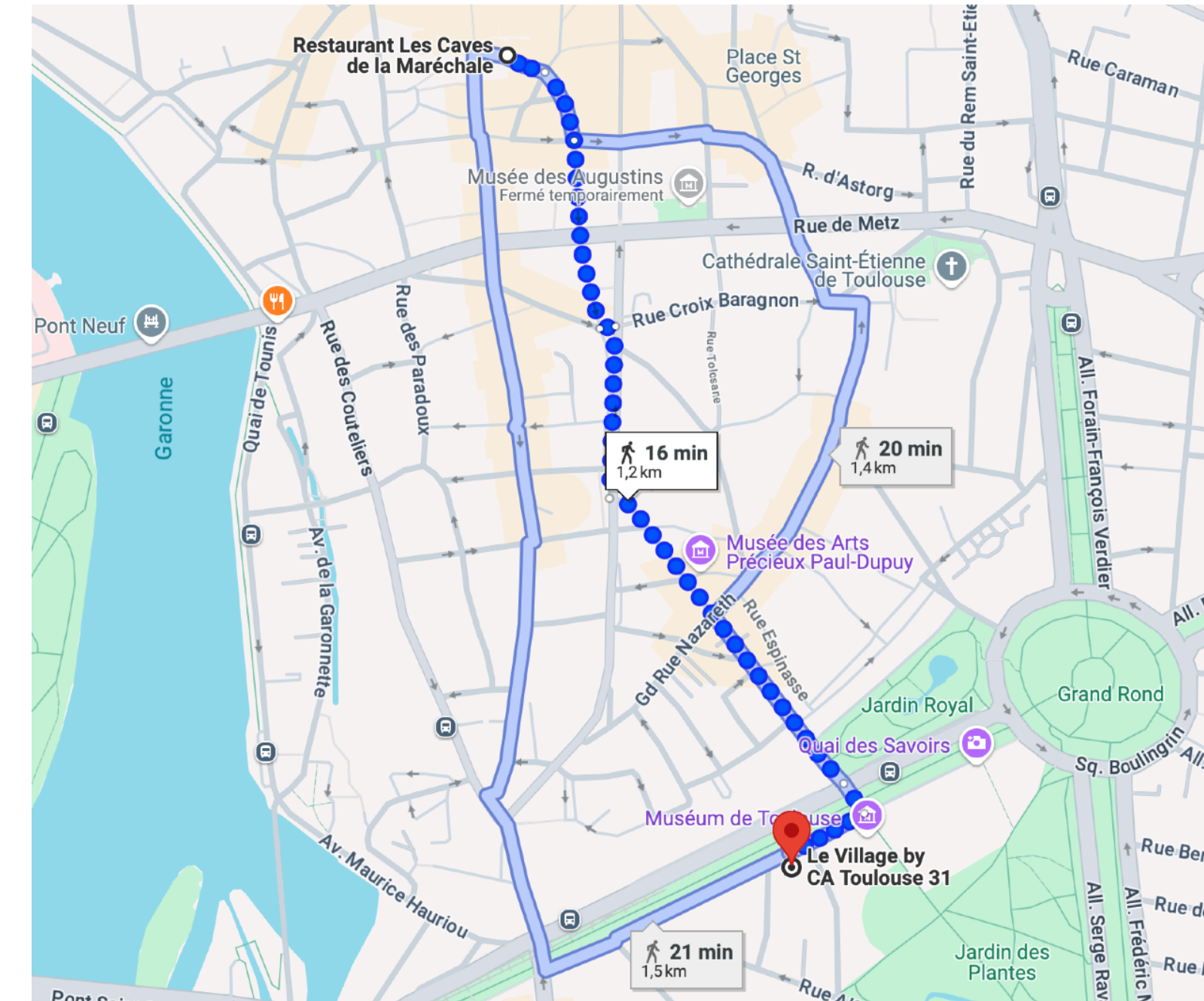
Workshop diner

Wednesday @19:30pm

Remember your choice ! ;-)

3 rue Jules Chalande 31000 Toulouse

Code of conduct

- This AISSAI Workshop is a scientific community event intended for presentations, networking and collaboration as well as learning. We value a civil and respectful environment which encourages the free expression and exchange of scientific ideas.

- This is an international event: please speak English if at least one non-French-speaking person is participating in the discussion.

- This is a multidisciplinary event: be open to other disciplines and to discussion.

AISSAI
AI for science, science for AI

If you have any problems or questions during the WS please contact us:
sos@l2it.in2p3.fr

Proposal for a workshop summary paper

We propose to write a workshop summary paper.

‣ More than a collection of individual results and contributions, aim for a concise synthesis, food for for thought.

‣ Finish before the end of the year 2024.

‣ Please talk to us / with everybody if you are interested.

‣ Jan volunteers to invest effort in editing and making this converge.

‣ This is a team effort - everybody who would like to contribute is welcome.

‣ Should be easy to find a journal - we know of journals that explicitly solicit such papers.

Let's keep this in mind and make a decision during the workshop.