



Línea 8. Computación, big data e inteligencia artificial

Overview





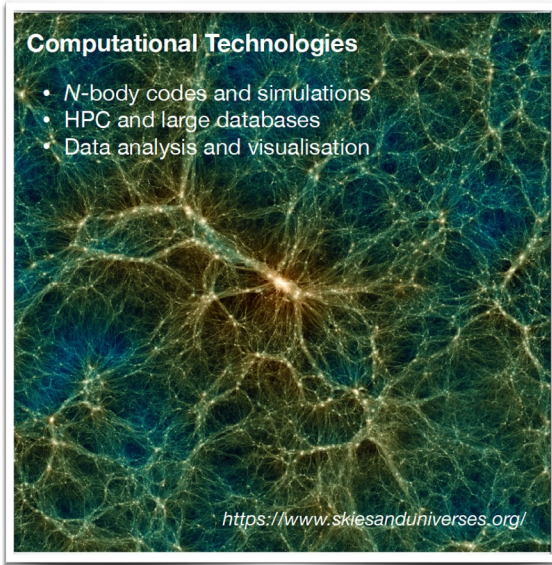
Andalucía

*Preparación del análisis de los datos masivos del cartografiado J-PAS en colaboración con Aragón. También se completará el desarrollo de un **prototipo de SKA Regional Center (SRC)**, una plataforma hardware, software y de servicios integrada en una red internacional. Este SRC apoyará el procesamiento y análisis de datos interferométricos y la reproducibilidad de los resultados científicos (principios FAIR). La implementación del SRC permitirá la capacitación de personal en áreas de impacto social como big data, data science o green computing, y constituirá un vehículo de colaboración y atracción de talento.*

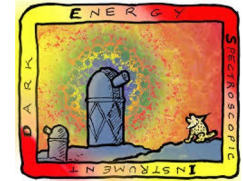
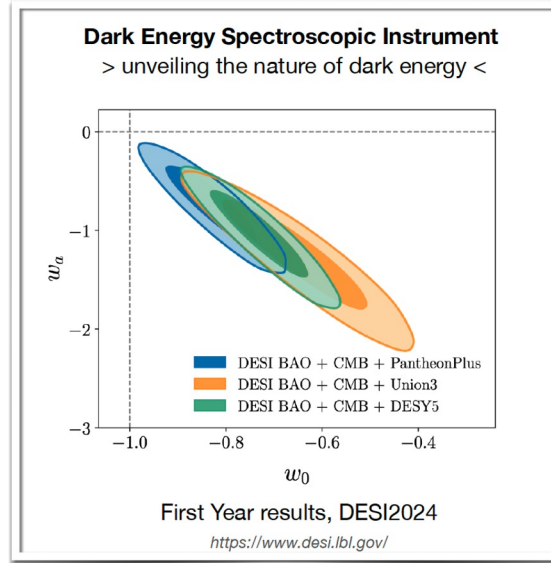


Andalucía

COSMOLOGY AND ASTROPARTICLE PHYSICS LAB



contact: f.prada@csic.es

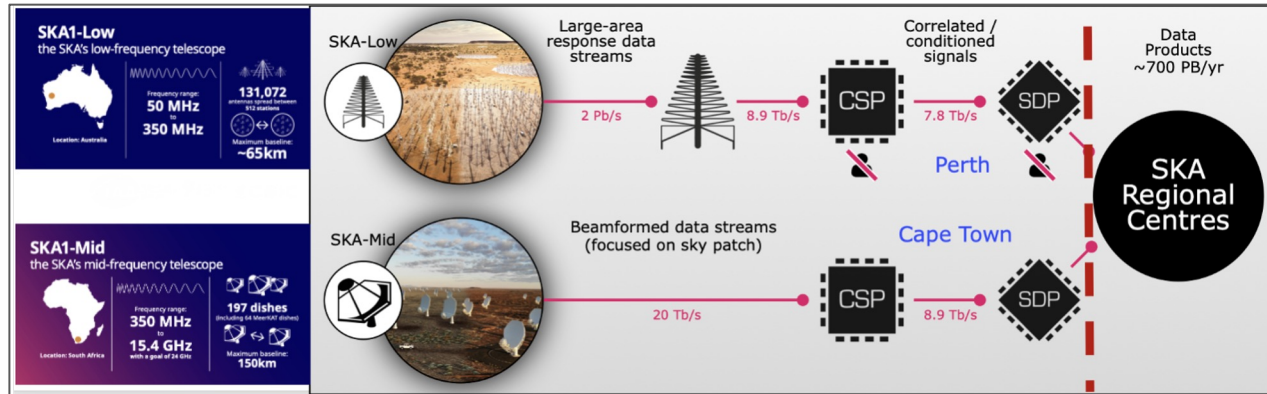


- Storage acquisition, 1 PB, for cosmological simulations database. **Tender in preparation.**

Andalucía

Fortalecimiento y actualización de la infraestructura informática para el nodo español del SKA Regional Centre

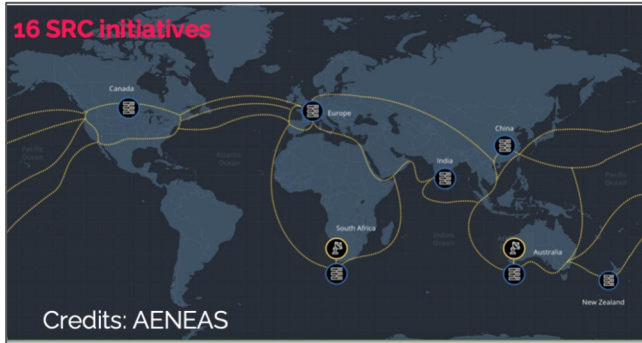
- Open key questions in Astrophysics, Astrobiology and Fundamental Physics
 - Formation of the 1st galaxies in a dark Universe dominated by atomic gas
 - Evolution of the atomic gas and star formation till the current epoch
 - Strong field tests of gravity using pulsars black holes
 - Active Galactic Nuclei and Galactic Centre
 - Extrasolar planets (protoplanetary disks, biomarkers)



Andalucía

Fortalecimiento y actualización de la infraestructura informática para el nodo español del SKA Regional Centre

- 16 SRC initiatives collaborating with SKAO to design the SRCNet
- SCRNet v0.1: The first operational version of the SRCNet, expected January 2025
- The Spanish SRC (espSRC) among the 9 EoI sent to SKAO to participate in SRCNet v0.1
- Hardware contribution to espSRC (SRCNet v0.1) will be acquired partially with project
- **espSRC computing infrastructure will be extended with:**
 - more than one Petabyte storage
 - +350 CPU cores and GPU computing



Current espSRC computing cluster

More details in Susana Sánchez's talk in this session



Aragón

*El cartografiado de galaxias J-PAS en el CEFCA producirá una gran cantidad de datos. Aragón jugará un papel determinante en las iniciativas relacionadas con la explotación de grandes volúmenes de datos producidos por grandes cartografiados astronómicos con fuerte implicación española, como J-PAS. Para dar apoyo a esta línea, se procederá a la **renovación parcial del equipo de la Unidad de Procesado y Archivo de Datos (UPAD)** del CEFCA, así como al desarrollo de software, big data y técnicas de inteligencia artificial.*

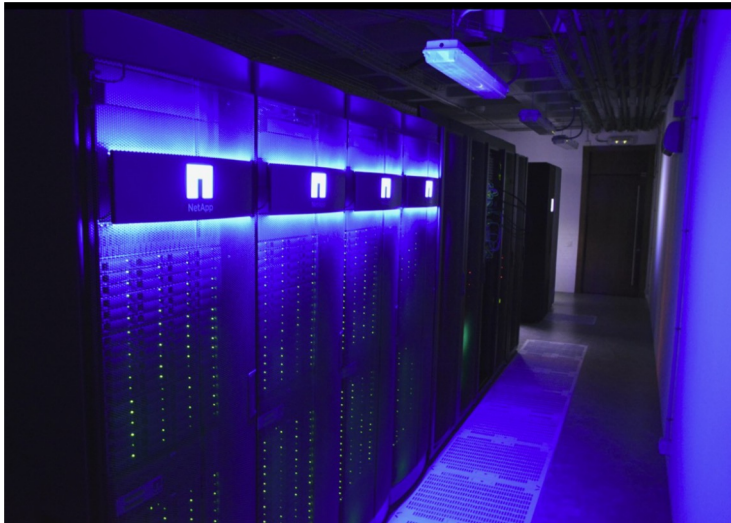
Aragón

- *Suministro, instalación y puesta en funcionamiento de EDAM*
 - **External Data Access Machine (EDAM):** It is the equipment that allows scientific data to be offered abroad, either within collaborations or projects, or to the public. **Action** implies renewal of this equipment, called *EDAM-NG (Next Generation)*.
 - **Specs:** 2 monitoring nodes and extended storage of ~100TB.
 - **Status:** Physical installation during 2023. Currently in the final phases of application installation and configuration. Imminent start-up.

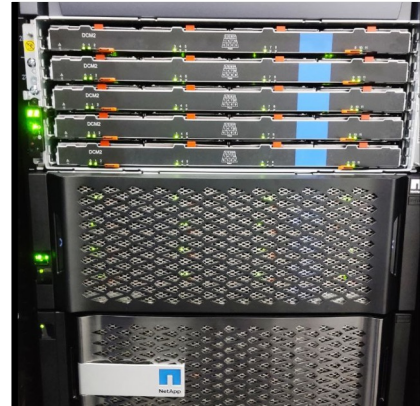


Aragón

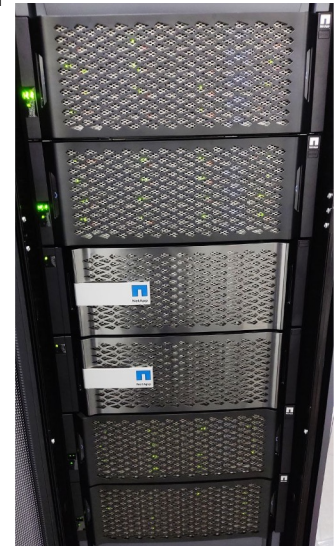
- *Suministro, instalación y puesta en funcionamiento de parte de la renovación del almacenamiento en disco de acceso rápido de la UPAD (CEFCA)*
 - **Tender result:** 1.2PiB @ 5000MB/s average read/write
 - **Status:** Implementation completed between May 27 and 29, 2024
 - **Pending:** Final acceptance, transfer of data from old to new system



Current UPAD storage to be renewed



Storage renewal purchased and implemented





Baleares

*Desarrollo de técnicas de computación avanzada para simulaciones y minado de datos aplicadas a ondas gravitacionales y astronomía multi-mensajero. En paralelo, se **adquirirá un superordenador regional** con soporte para aprendizaje automático de alto rendimiento dedicado a la astrofísica computacional, en las áreas de ondas gravitacionales y física solar, y se apoyará a los grupos científicos que lo explotarán.*



Baleares

- Performing numerical simulations of black hole mergers.
- Development and optimization of data analysis tools to search for gravitational signals.
- Contribution to several LIGO-Virgo-KAGRA publications of results derived from the third observation period.
- Acceleration of blind searches for gravitational waves through the use of GPUs
- Launch of a [Kaggle competition](#) for data analysis of continuous gravitational signals with artificial intelligence techniques:
- Development of machine learning methods and new implementations of classic algorithms to obtain results with greater robustness and efficiency compared to previous implementations.
- Procedures have begun to define the characteristics of a regional supercomputer with support for high-performance machine learning dedicated to computational astrophysics.



Software Contribution:

- David Keitel *mantenedor de los códigos de análisis de datos de ondas continuas ("LALPulsar") en el paquete público de código abierto LALSuite*. DOI: [10.7935/GT1W-FZ16](https://doi.org/10.7935/GT1W-FZ16).
- Anna Heffernan. *Self-Force Regularisation Parameters Package (Mathematica)*. DOI: [zenodo.6282572](https://doi.org/zenodo.6282572)
- Tenorio, Rodrigo; Modafferi, Luana M.; Keitel, David; Sintés, Alicia M. *distromax*. DOI: [zenodo.5763765](https://doi.org/zenodo.5763765) A Python package to empirically estimate the loudest candidate from a gravitational-wave search.
- Ashton, Gregory; Keitel, David; Prix, Reinhard; Tenorio, Rodrigo, *PyFstat*. DOI: [zenodo.3967045](https://doi.org/zenodo.3967045) A python package for gravitational wave analysis with the F-statistic.
- Alicia M. Sintés, S. Husa, D. Keitel, M. Colleoni, H. Estellés, C. García-Quirós, A. Ramos-Buades, R. Tenorio and others (contributors). *LALSuite - LSC Algorithm Library Suite*. DOI: [10.7935/GT1W-FZ16](https://doi.org/10.7935/GT1W-FZ16)

Baleares

Mentions and awards

- Joan Llobera Querol y Lluç Planas, 1st position in the G2Net Cost Action School Hackathon: 4th Training School on Gravitational Waves, Geophysics and Machine Learning in Thessaloniki, Greece. (<https://indico.physics.auth.gr/event/14/>) 28–31/03/ 2023
- René Mérou Mestre awarded with accésit at **VIII Premios TalentIC**, January 2024
- Rodrigo Tenorio became liason for working groups *Continuous gravitational waves* (CW) and *Artificial Intelligence* (MLA) for the collaboration LIGO-Virgo-KAGRA.

Recent Supercomputing Resources:

- RES-BSC Pirineus (CSUC): "*Preparations for a catalog of generic gravitational wave signals from binary black hole coalescence*". Ref: AECT-2024-1. IP: Sascha Husa. Importe: 2.635kh. Fecha: 01/03/2024 - 30/06/2024.
- RES-BSC Picasso (UMA): "*Robust parameter estimation and lensing studies with the LVK network at O4 sensitivity and beyond*". Ref: AECT-2023-3-0020. IP: David Keitel. Importe: 1400kh. Fecha: 01/11/2023 - 29/02/2024
- RES-BSC Marenostrum: "*Preparations for a catalog of generic gravitational wave signals from binary black hole coalescence*". Ref: AECT-2023-3-0022. IP: Sascha Husa. Importe: 5.300kh. Fecha: 01/11/2023 - 29/02/2024
- Artemisa call 2024.1: "*Accelerating the search for continuous gravitational-wave signals from unknown millisecond neutron stars in binaries*". IP: Rodrigo Tenorio, David Keitel. Fecha: 01/03/2024 - 30/06/2024

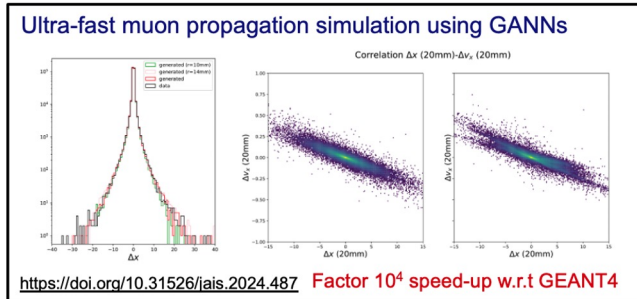


Cantabria

*Desarrollo de herramientas de big data y machine learning para su aplicación al procesamiento y análisis de los datos reales y simulados producidos en J-PAS, Euclid, SKA, QUIJOTE, LiteBIRD y Athena. Aplicación de **técnicas de machine learning al análisis de datos del LHC**: integración de GPU y FGPA en la Grid; aplicación al trigger de estos procesadores, entre otras aplicaciones.*

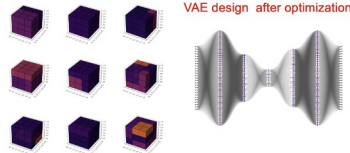
Cantabria

- Development of machine learning tools and machine learning techniques for LHC data analysis



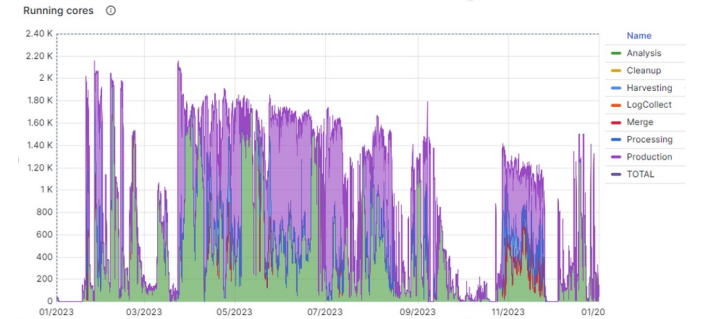
Application of ML techniques

Anomaly detection using Variational Auto-Encoders in the context of muon tomography applied to cargo inspection



Regularity patterns (examples)

- System trained with geometrical patterns with density regularity (boxes, pallets, etc)
- Triggering alarm for abnormal cases



CMS infrastructure for the LHC



- In the process of hiring a computer scientist.

More information in Miguel Villaplana's presentation in this session.



Catalunya

Acceso eficiente a **grandes conjuntos de datos astrofísicos (CosmoHub)**, hub de astronomía multi-mensajero. Además, fortalecer el equipo humano responsable del desarrollo de herramientas para la gestión de alertas y planificación (scheduling). Más generalmente, se reforzarán los equipos que trabajan en la generación, gestión y **explotación de datos masivos** de los experimentos descritos anteriormente, consolidando y expandiendo su know-how en las herramientas necesarias para hacer efectiva la data science, así como desarrollos de computación avanzada para la descripción de fenómenos astrofísicos complejos.



CSIC

Institute of
Space Sciences

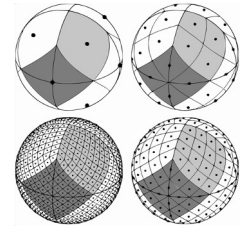
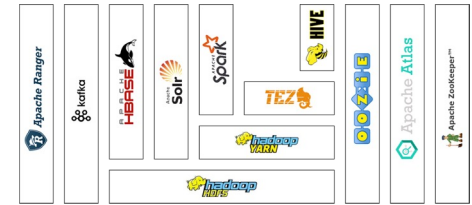


Institut de Ciències del Cosmos
UNIVERSITAT DE BARCELONA

Catalunya



- Development of a **comprehensive Multi-Messenger (MM) Data Hub** (expansion of the CosmoHub service)
 - Compile and organize **hundreds of massive datasets** from multiple messengers (optical, gammas, GWs and neutrinos)
 - Provide an easy and intuitive interface based on standard protocols
 - **Short access latency** and the ability to work with multi-terabyte datasets with ease
- Runs on top of our **Big data common service** (advanced distributed architecture):
 - 4th generation of the hardware platform that powers this service.
 - **Customized Hadoop distribution** (tested and deployed using CI/CD pipelines)
- MM platform *particular* features:
 - Standard data formats (CSV, Parquet, FITS, ASDF, DL3)
 - Three sets of **User Defined Functions**: HEALPix, Array Aggregation, Spherical geometry + MOC maps
 - Support to multiple Virtual Observatory (VO) protocols (TAP, **ADQL**, VOTable and UWS)
 - Supplementary storage and computing resources
- Personnel already hired:
 - Two Postdoc researchers (Development Operations and Data science and data analysis)
 - Senior Technician (Operations)
 - Predoc researcher (Data science)

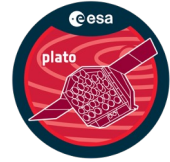


More details in Jorge Carretero's talk in this session

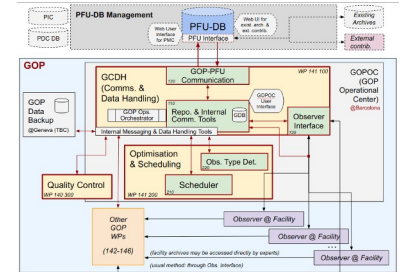
Catalunya

- Development of stellar model libraries to determine fundamental parameters of the 150k stars of the PLATO Core Program, a component of the MSAP5 pipeline
- IA developments for PLATO
- Computing cluster
- Development of the scheduler software for PLATO earth observations: scheduler and graphical visualization interface (in close collaboration with ICC-UB, IEEC & ICE)
- **Hired software engineer and post-doctoral researcher**

- Definition of the data management system for the Ground-based Observing Program (GOP) for PLATO. Responsible for various work packages together with ICE
- Design, implementation and testing of GOP management and monitoring subsystems
- **Hired software engineer**
- **Definition and purchase of the GOP operations center (calculation and storage server)**



CSIC Institute of Space Sciences



Institut de Ciències del Cosmos
UNIVERSITAT DE BARCELONA



Comunidad Valenciana

Con su nodo local de la **RES y ARTEMISA**, la comunidad valenciana cuenta con una infraestructura de cálculo, basada en CPUs y GPUs, orientada a la supercomputación, el aprendizaje profundo y la inteligencia artificial. El programa permitirá su actualización y explotación, potenciando el grupo humano dedicado al desarrollo de algoritmos. También se fortalecerán tanto el equipo humano como el equipamiento para la **simulación, gestión y explotación de datos masivos de los experimentos en colisionadores** de las próximas décadas, empezando con el HL-LHC. Asimismo, se potenciará el desarrollo de la astrofísica y cosmología computacional, en colaboración con Baleares.



Comunidad Valenciana



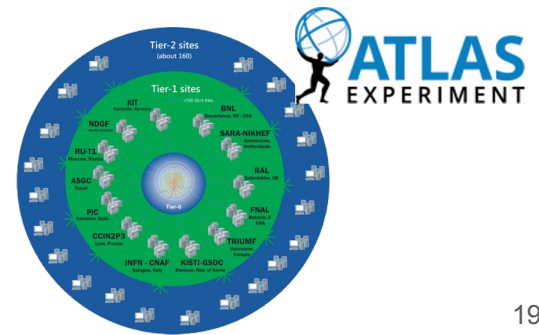
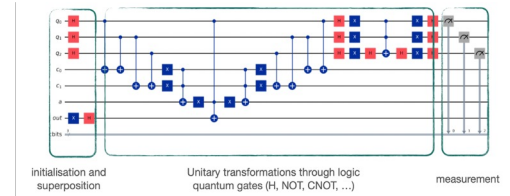
*ARTificial Environment for ML and Innovation
in Scientific Advanced Computing*

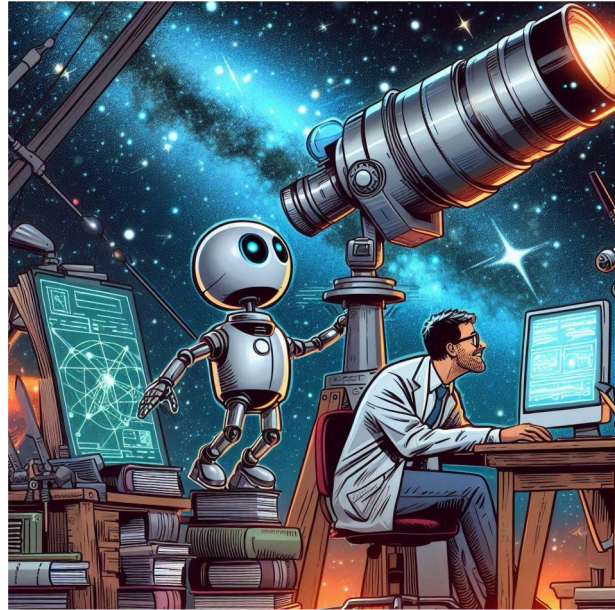
- **Artemisa:** GPU-intensive computing infrastructure dedicated to Artificial Intelligence(AI) and Machine Learning (ML) located at IFIC's data centre. Its advanced features and excellent performance enable the accelerated development of projects involving AI areas.
- Capable of ML tasks involving large amounts of data to produce empirical models in physics, chemistry, biology and social studies.
- Provides last generation GPUs, even featuring a 8-GPU A100 server.
- Steadily increase of projects using Artemisa.
 - Artemisa has achieved high engagement and growth rate
 - GPU processing hours have increased from machine learning tasks and, particularly, deep learning tasks.
- Ongoing:
 - **Documentation and user experience**
 - **Infrastructure general developments:** Improve system resource monitoring. Keep the infrastructure up-to-date adding servers with state of the art GPUs, even in a high-demand market.
 - **Upgrading the infrastructure.** New servers dedicated to Artemisa online usage (Jupyterhub or similar)
 - **New computer scientist for development and support. Acquisition of new equipment in progress.**



Comunidad Valenciana

- More on infrastructure:
 - Provided additional equipment for parallel and serial computing for research projects:
 - *"Ondas GRavitacionales, Axiones y Materia Oscura, Lattice e Inteligencia Artificial"* y *"Algoritmos cuánticos en fenomenología de partículas elementales"*
 - Collider Experiments:
 - ATLAS Tier2 experiment contribution with storage and processing machines
 - Further development of the computing infrastructure for next years' pledges (HL-LHC)
 - Replacing decaying infrastructure and future points of failure in storage
 - Evolution of the Analysis Facilities for more interactive and user friendly use
 - Post-doc and engineer hired for support and development of the infrastructures
- Development of exploitation of massive data from experiments with colliders, astrophysics and computational cosmology with ML/AI techniques (more in the plenaries and parallel talks).





Credits to Copilot

