Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
ICSC Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

ASTRAI

Advanced Supernova Transient Research with Artificial Intelligence

Luca Naso, Koexai S.r.l.

Spoke 3 Progetti Bandi a Cascata, 24/09, 2024

ASTRAI

Advanced Supernova Transient Research with Artificial Intelligence

Tematica: 6

Metodologie Avanzate di Analisi Dati

Sotto-tematica: c)

Tecniche numeriche e codici basati sull'apprendimento automatico al fine di modellare fenomeni esplosivi come Supernovae e altri fenomeni transitori simili

Project Overview: challenge and proposed solution

Supernovae (SNe) play a crucial role in various astrophysical processes, including nucleosynthesis, dust production, cosmic ray acceleration, and the emission of neutrinos and gravitational waves. Today, scientists study SNe using complex statistical methods, like Gaussian and Bayesian Processes, to reconstruct light and velocity curves and then infer the progenitor main physical characteristics. This process is **time-consuming**, often taking hours for a single curve, which slows down the overall pace of scientific research. Moreover, the upcoming Legacy Survey of Space and Time (**LSST**) is expected to generate a massive amount of observational data that will need to be analyzed quickly.

To address these challenges, ASTRAI proposes the use of **Machine Learning** (ML) techniques. While ML holds great promise, it requires large training datasets. Currently, there are no sufficiently large datasets of SNe available. To overcome this limitation, ASTRAI will use **Generative AI** techniques to create a synthetic dataset, thus lowering the number of models that needs to be simulated with the current statistical methods. The large synthetic dataset can then be used as a training dataset for an **automatic data analysis system**, significantly accelerating the pace of SNe research.

Technical Objectives and Methodologies

Astrophysical Characterization of SNe: development of an innovative analytical model for the analysis of light curves of hydrogen-rich SNe, including various post-explosion heating mechanisms. Application of Bayesian statistical methods for the characterization of light curves and estimation of fundamental physical parameters (e.g. Energy, mass, ...).

Development of synthetic observations using Generative AI (GenAI): exploration of GenAI algorithms, including Variational Autoencoders, Stable Diffusion models, and Generative Adversarial Networks (GANs), for the generation of a comprehensive dataset utilizing existing observations and simulations from astrophysical models.

Automatic Characterization of SNe: application of cutting-edge AI algorithms for the automatic characterization of SN events. Examination of algorithms such as Transformers, RNN-LSTM, Random Forest, and SVM with a focus on fine-tuning promising models to maximize accuracy and reliability. Training performed on the synthetic dataset.

Computing Resources

Astrophysical Characterization of SNe:
Standard approach requires to simulate 5e9 models, producing outputs for about 100 TB, requiring about 80k hours (9 yr) on single machine (8 CPU, 16 GB RAM), for low interacting H-rich SNe

Synthetic Observations creation (GenAI) & Automatic Characterization of SNe (ML):
Proceed with trial and error, testing several approaches and neural network architectures on smaller dataset, and then training the final model on a larger dataset.
Expected
1600 hours spread across different machines (from 16 to 192 vCPUs, from 64 to 728 GiB Memory, with NVIDIA A10G Tensor Core GPU, from 35 to 280 TFLOPs), cost on AWS ~6k $

Involved Staff and new recruitments

Spoke Scientific Advisor: Maria Letizia Pumo, UniCT

Koexai Team size: 1.66 FTE = 2858 hours

1. **Luca Naso:** Project Lead, Senior Data Scientist, Ph.D. in Astrophysics, and CEO of Koexai
2. **Marco Cataldo:** Project Manager and Project Administrative Manager
3. **Vincenzo Del Zoppo:** Senior Data Scientist and AI Engineer
4. **Giuseppe Puglisi:** Data Scientist, Ph.D. in Machine Learning
5. **Fabio Spampinato:** Data Scientist
6. **Andrea Claudio Grasso:** Data Scientist

Scientific Consultant:

- **Stefano Pio Cosentino:** Ph.D. candidate in Astrophysics

Gantt

OR	AR	Titolo	T1			T2			T3			T4		
			1	2	3	4	5	6	7	8	9	10	11	12
1		*Studio e definizione di modelli fisici per la caratterizzazione astrofisica degli eventi di SN*	■	■	■	■	■	■	■	■	■	■	■	■
	1.1	Studio dello stato dell'arte	■	■										
	1.2	Studio e definizione del modello analitico		■	■	■	■	■	■	■	■	■		
	1.3	Stima dei parametri fisici				■	■	■	■	■	■	■	■	■
2		*Studio e definizione di un framework di Intelligenza Artificiale per la generazione di dataset sintetici di osservabili da eventi di SN*	■	■	■	■	■	■	■	■	■	■	■	■
	2.1	Studio, ricerca e analisi delle osservazioni astronomiche disponibili	■	■										
	2.2	Studio e definizione del modello di AI generativa per la creazione dei dataset			■	■	■	■						
	2.3	Evoluzione e determinazione del modello di AI tramite confronto con i modelli astrofisici							■	■	■	■	■	■
	2.4	Creazione e condivisione di un ampio dataset di osservabili							■	■	■	■	■	■
3		*Studio e definizione di modelli di Machine Learning avanzati per la caratterizzazione di eventi di SN*						■	■	■	■	■	■	■
	3.1	Caratterizzazione tramite algoritmi avanzati di AI						■	■	■	■			
	3.2	Ottimizzazione degli algoritmi avanzati									■	■	■	■

Timescale, Milestones, SAL

Project start: Sep/24

Milestone 1 - Nov/24
- Literature Study
- 1st Analytical model
- AR1.1
- AR2.1

Milestone 2 - Feb/25
- Gen AI model
- AR2.2

Milestone 3 - May/25
- Building the dataset
- Automatic characterization system
- AR3.1

Milestone 4 - Aug/25
- End of project: all deliverables
- AR1.2, AR1.3
- AR2.3, AR2.4
- AR3.2