Explainable autoencoder for neutron star dense matter parameter estimation

Francesco Di Clemente[1,2], **Matteo Scialpi**[2,3], Michał Bejger[2,4]

ET Symposium 2025

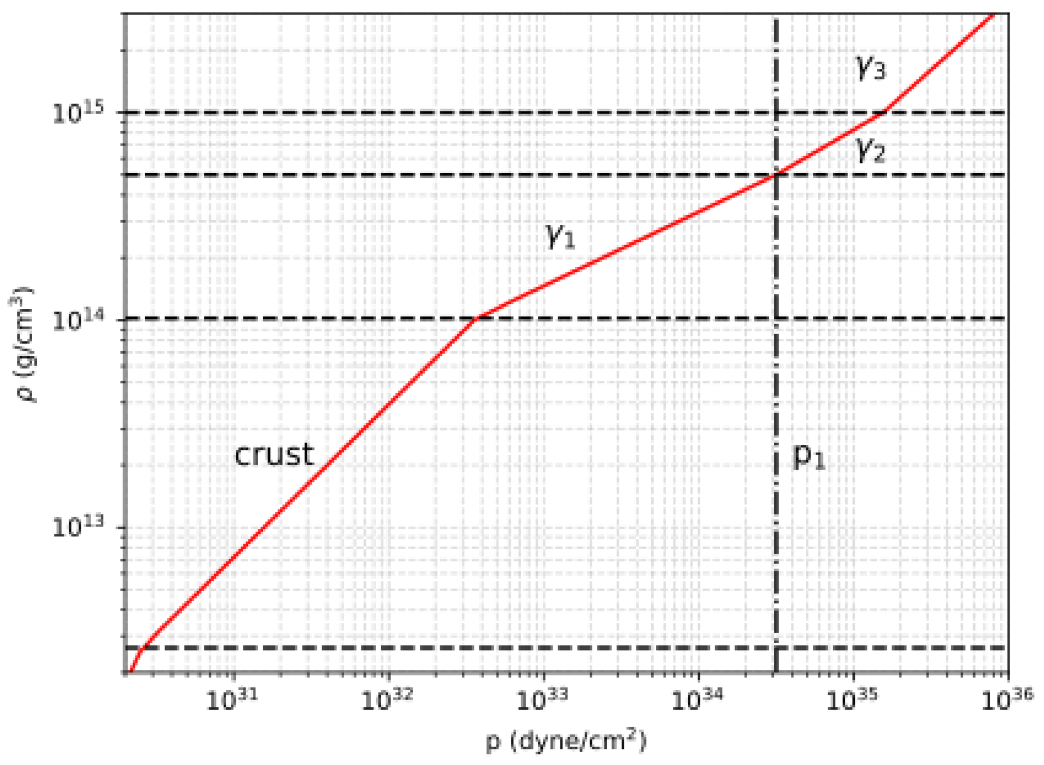
[1]Department of Physics, University of Houston, Houston, TX 77204, USA
[2]INFN Sezione di Ferrara, Via Saragat 1, 44122 Ferrara, Italy
[3]Dipartimento di Fisica e Scienze della Terra, Università di Ferrara, Via Saragat 1, 44122 Ferrara, Italy
[4]Nicolaus Copernicus Astronomical Center, Polish Academy of Sciences, Bartycka 18, 00-716 Warszawa, Poland

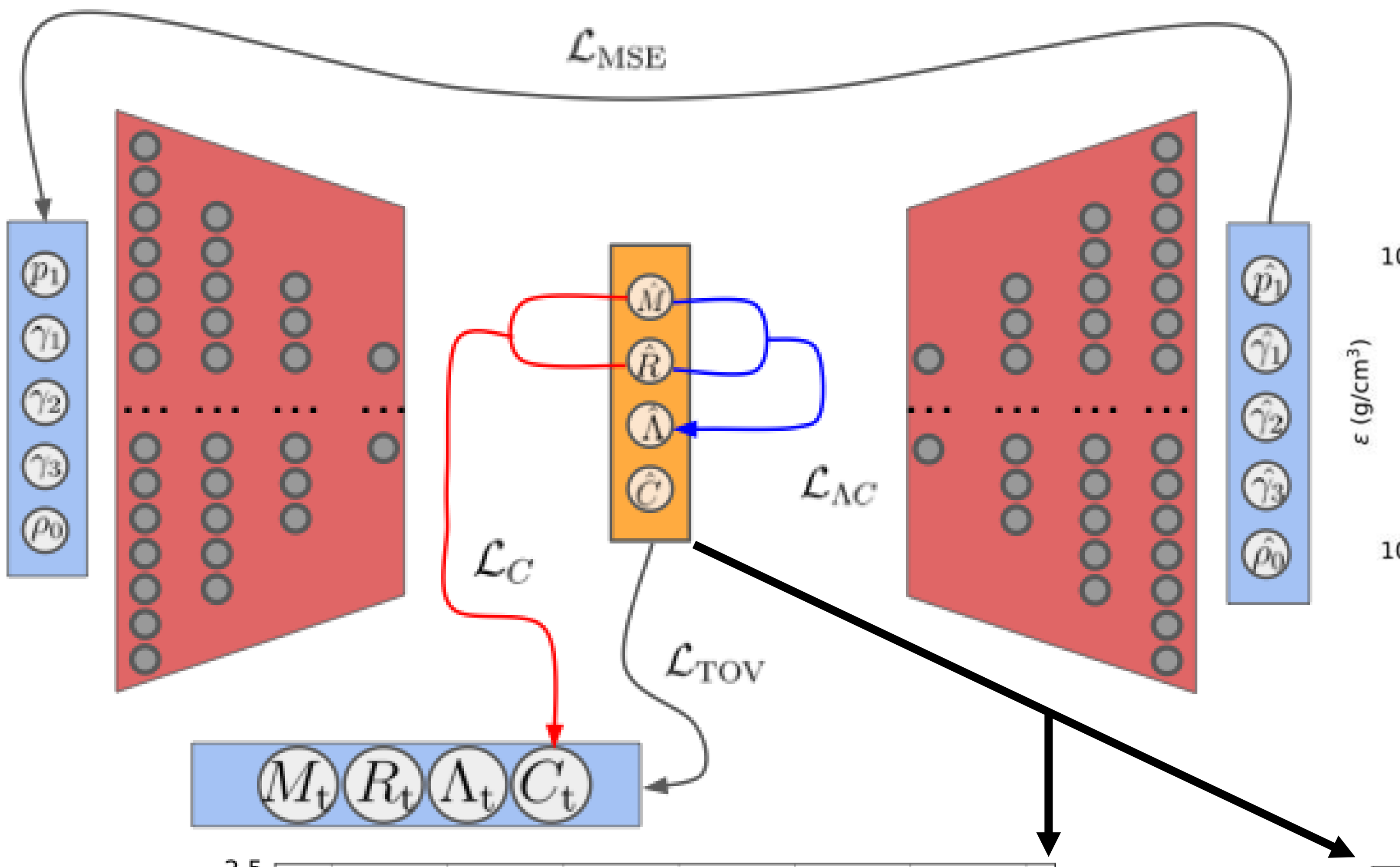
INFN Istituto Nazionale di Fisica Nucleare

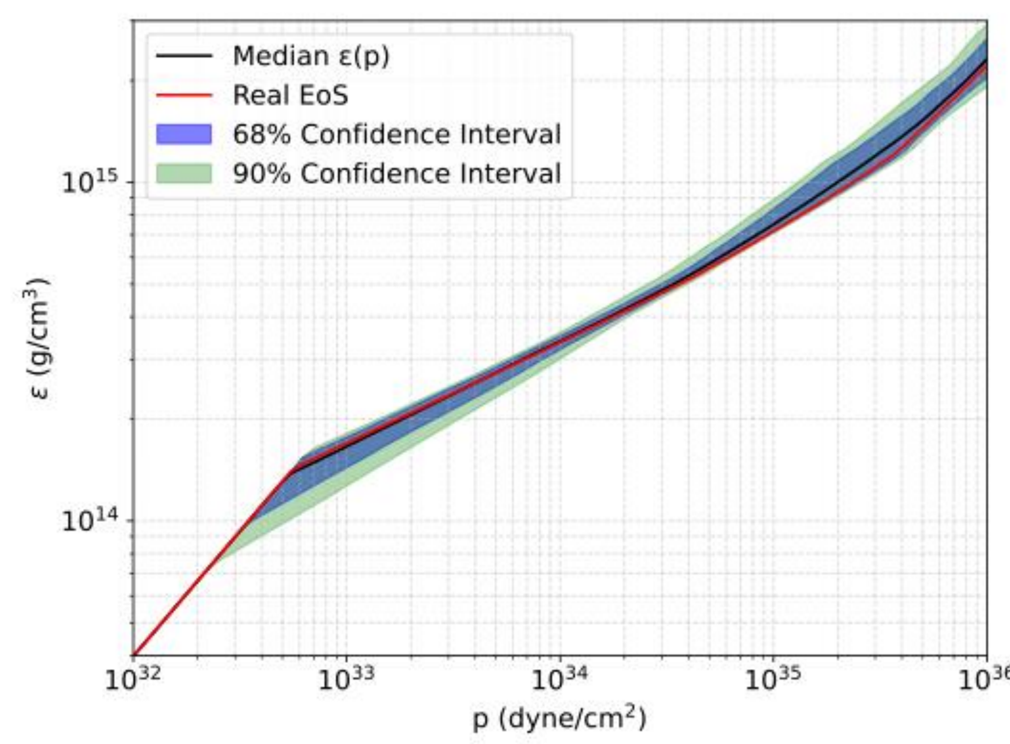
Ferrariae Universitas · Ex labore fructus · 1391

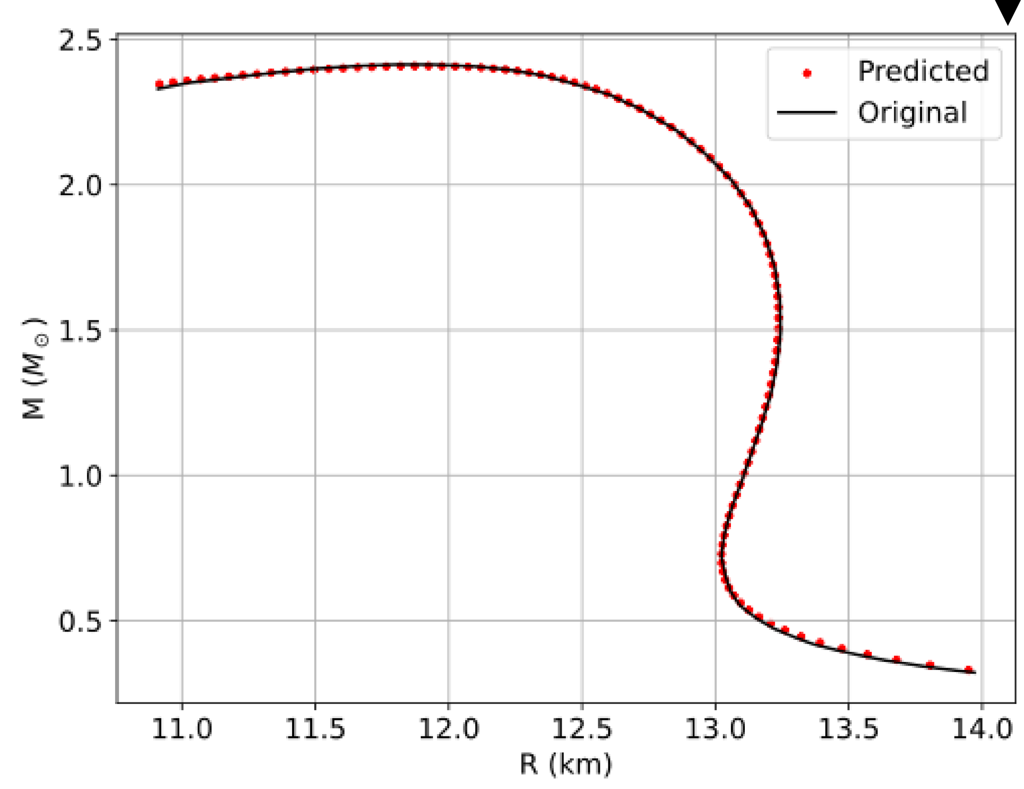
Abstract

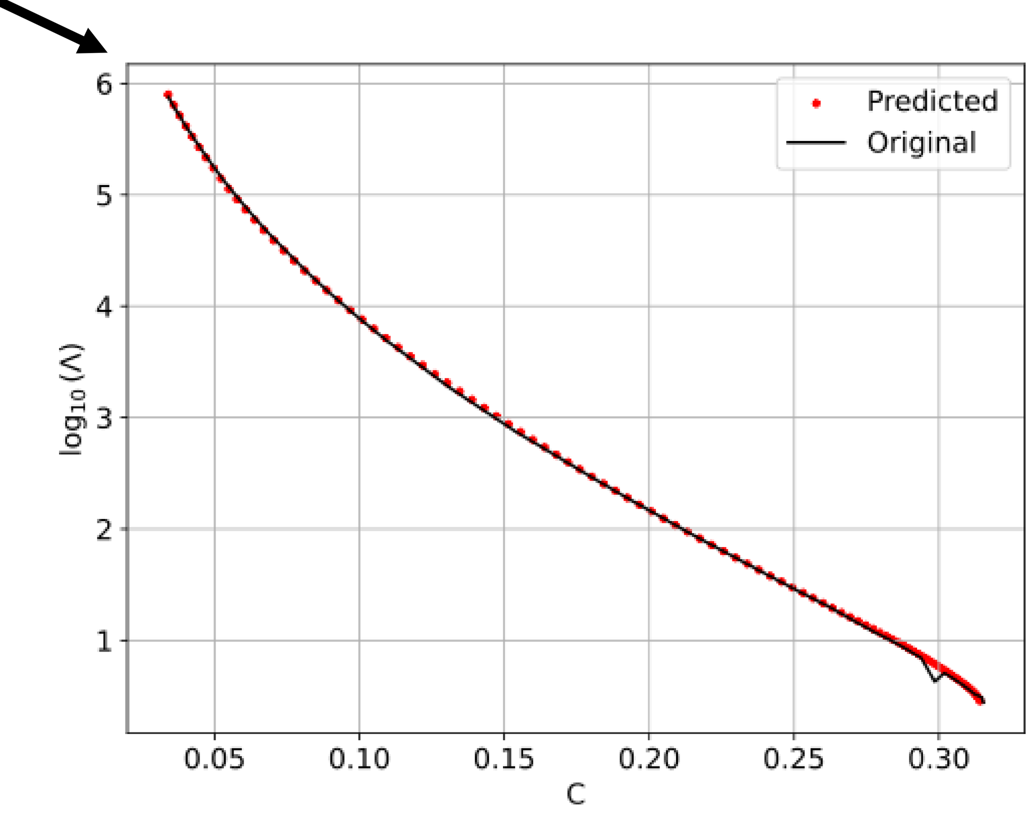
In this poster we present a Physics-Informed Autoencoder (PIA) designed to encode the equation of state of neutron stars into an interpretable latent space. The input polytropic EoS is encoded in the mass, radius, and tidal deformability values of a neutron star. Unlike traditional black-box autoencoders, our approach incorporates additional loss functions to enforce explainability in the encoded representations. This method enhances the transparency of machine learning models in physics, providing a robust proof-of-concept tool to study compact stars data. We present our results, which demonstrate that the proposed autoencoder not only accurately estimates the EoS parameters and central density/pressure but also offers insights into the physical connection between equation of state and observable physical quantities. We also discuss implications for ET, thanks to which neutron-star physics will be largely enriched by new GW observations.











1 Introduction

Goals:

1. To solve the TOV equations for neutron star dense matter (from EoS to macroscopic quantities).
2. To integrate the TOV solution (from macroscopic quantities to EoS).

- **Piecewise-polytropic EoS** (left plot):

$$p(\rho)=\begin{cases}K_1\rho^{\gamma_1}, & \text{if } \rho\le\rho_1,\\ K_2\rho^{\gamma_2}, & \text{if } \rho_1<\rho\le\rho_2,\\ K_3\rho^{\gamma_3}, & \text{if } \rho>\rho_2,\end{cases} \quad \text{with} \quad \begin{aligned}K_1&=\rho_1^{-\gamma_1},\\ K_2&=K_1\rho_1^{\gamma_1-\gamma_2},\\ K_3&=K_2\rho_2^{\gamma_2-\gamma_3},\end{aligned}$$

p_1: pressure at the joint between ρ_1 and ρ_2;
ρ_0: density at the joint between crust and ρ_1.

- **Macroscopic quantities**: mass M, radius R, tidal deformability Λ and compactness C of the neutron star.

2 Autoencoder and Loss

We use an autoencoder (upper central figure) **to recover the input EoS parameters passing by an interpretable latent space.**

- The encoder solves the TOV, the decoder integrates the solution.
- This can be done thanks to **physics-informed loss**:

$$\begin{aligned}\mathcal{L} =&\mathcal{L}_{\rm MSE}+\mathcal{L}_{\rm TOV}+\mathcal{L}_{\rm C}+\mathcal{L}_{\Lambda C}+\mathcal{L}_{\rm b}\\ =&\|(\hat{\mathbf{p}}_1,\hat\gamma_1,\hat\gamma_2,\hat\gamma_3,\hat\rho_0)-(\mathbf{p}_1,\gamma_1,\gamma_2,\gamma_3,\rho_0)_{\rm t}\|^2+\left\|(\hat M,\hat R,\hat\Lambda,\hat C)-(M,R,\Lambda,C)_{\rm t}\right\|^2+\\ &+\left\|\frac{\hat M}{\hat R}-C_{\rm t}\right\|^2+\left\|\hat\Lambda-f\left(\frac{\hat M}{\hat R}\right)\right\|^2+\sum_i\left[{\rm ReLU}^2(\gamma_{\rm i,mins}-\gamma_{\rm output})+{\rm ReLU}^2(\gamma_{\rm output}-\gamma_{\rm i,maxs})\right]\end{aligned}$$

- **Training dataset:** 10^4 piecewise-polytropic EoS realizations (p_1, γ_1, γ_2, γ_3, ρ_0) and their target TOV solution for M, R, Λ and C.

3 Results

The autoencoder recovers both the TOV soution and the EoS parameters.

- In the lower central and lower right plots one can see the recovered M, R, Λ and C against their true value for DD2 EoS case.
- In the upper right plot one can see the recovered EoS for DD2 case. The confidence intervals are present because of the active Monte Carlo dropout in the decoder and is consistent with the micro-physics uncertainty when one (M,R,Λ,C) point is given.

4 Discussion

Goals were reached:

- The parameters of the EoS are well recovered, passing through the constraint latent space with the macroscopic quantities.

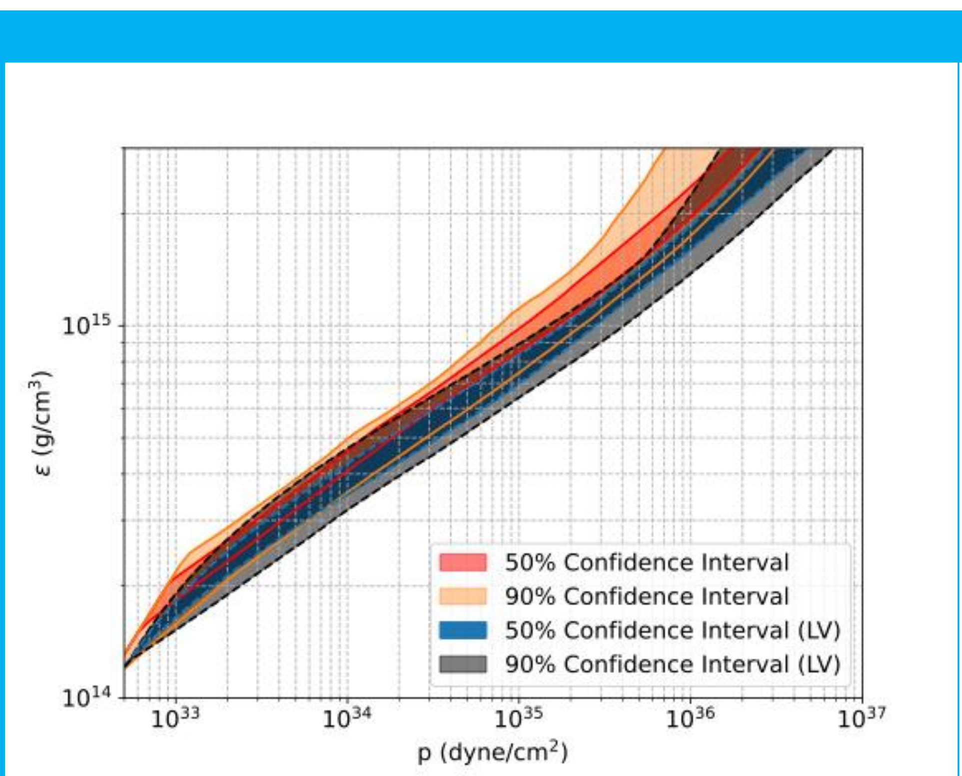
The autoencoder offers useful insights on the physical link between micro- and macro-physics:

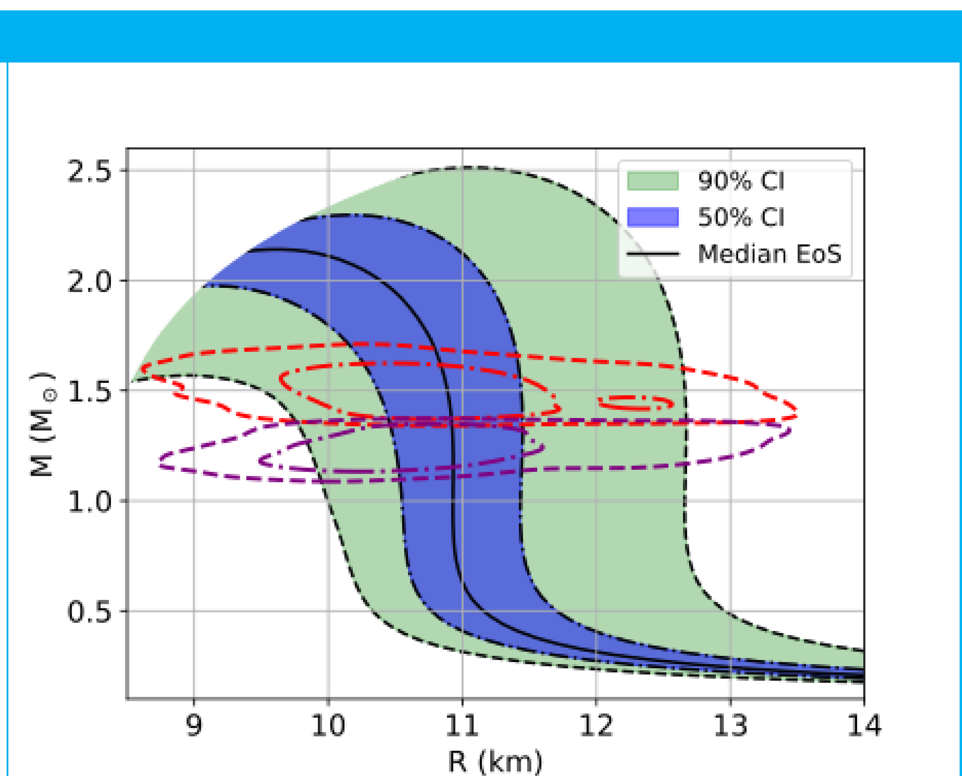
- We applied the best trained model to GW170817 and gave our prediction for the EoS of the two coalescing NSs.
- In ET the MR plot will be extremely more populated of events, so we will have more inputs to adjust our prediction.
  - In LVK we already have more than one BNS merge, but since no electromagnetic counterpart was observed, the needed MR samples are too sparse.

GW170817

We apply our trained decoder, and then our tarined encoder, to GW170817 (M,R Λ,C) data.

- In the left plot one can see the EoS results compared to the LV one. Notice that the results are compatible, but on the upper extrapolation part.
- In the right plot one can see our prediction for the MR representation of the recovered EoS for the two coalescing neutron stars.





Reference paper

arXiv

IOP – Machine Learing Science and Technology

https://arxiv.org/abs/2501.15222

https://iopscience.iop.org/article/10.1088/2632-2153/add3bd

Ministero dell'Università e della Ricerca

Italiadomani — Piano Nazionale di Ripresa e Resilienza

Finanziato dall'Unione europea NextGenerationEU

Funded by the European Union-Next Generation EU, Mission 4 Component 1 CUP J53D23001550006 with the PRIN Project No. 202275HT58