BART-Lagrangian [BART-L]

When Transformers Write Particle Physics Lagrangians

Yong Sheng Koay
Rikard Enberg
Stefano Moretti
Eliel Camargo-Molina

Department of Physics and Astronomy,
Uppsala University,
Box 516, SE-751 20
Uppsala, Sweden

UPPSALA UNIVERSITET

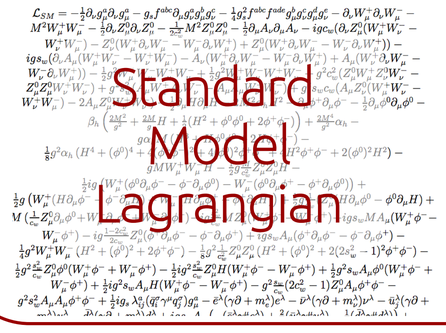
Based on work done in [1]

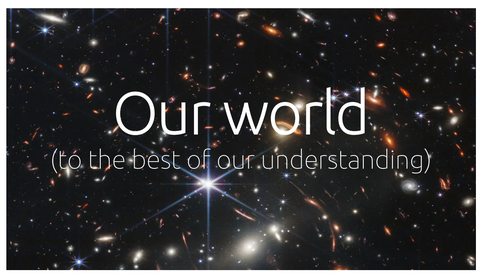
Can Transformers write Lagrangians?

YES!

Try it out yourself!

In particle physics, the Lagrangian serves as the fundamental mathematical object from which all interactions and predictions of a theory are derived. For example:





How does it work?

Natural Langauge : Never Gonna Give You Up ...

Lagrangians : $D_\mu \Phi^\dagger D^\mu \Phi + m^2 \Phi^\dagger \Phi + \lambda \Phi^\dagger \Phi \Phi^\dagger \Phi + \lambda' \Phi \Phi'' \Phi''' + i \Psi^\dagger_L \bar{\sigma}^\mu D_\mu \Psi_L \ldots$

Data Generation

Given the symmetries and particle content, one can write the symmetry conserving Lagrangian

Fields → AutoEFT [2] → Lagrangians

Model Task

Transformers works great with natural language, it should work great for a physicist's Language too!

Tokenized Fields → BART-L → Tokenized Lagrangians

Bidirectional and Auto-Regressive Transformer - Lagrangians

Example Tokenization Scheme

Object	Tokens
$\phi(\mathbf{3}, \mathbf{2}, -1/3)$	FIELD, SPIN, 0, SU3, -, 3, SU2, 2, U1, -, 1, /, 3, ID3
$\psi_L(\bar{\mathbf{1}}, \mathbf{3}, 0)$	FIELD, SPIN, 1, /, 2, SU2, 3, HEL, 1, /, 2, ID9
$\phi^\dagger(\mathbf{1}, \mathbf{1}, 7/5)$	FIELD, SPIN, 0, U1, 7, /, 5, DAGGER, ID7
$D_\mu^{(SU(2),U(1))}$	DERIVATIVE, SU2, U1, ID4
$D_\mu^{(SU(3),SU(2),U(1))}$	DERIVATIVE, SU3, SU2, U1, ID1
∂_μ	DERIVATIVE, ID5
$\bar{\sigma}^\mu$	SIGMA_BAR, ID2
$g^{\mu\nu}$	LORENTZ, ID6, ID4
ϵ^{ab}	SU2, ID0, ID7

Term : $\lambda\, \phi\, \phi^\dagger\, \Phi\, \Phi^\dagger$

Information	Tokens
λ	
$\phi(\mathbf{3}, \mathbf{2}, 6)$	FIELD, SPIN, 0, SU3, 3, SU2, 3, U1, 6, ID5
$\phi^\dagger(\bar{\mathbf{3}}, \mathbf{2}, -6)$	FIELD, SPIN, 0, SU3, -, 3, SU2, 3, U1, -, 6, DAGGER, ID7
$\Phi(\bar{\mathbf{3}}, \mathbf{1}, -7/2)$	FIELD, SPIN, 0, SU3, -, 3, U1, -, 7, /, 2, ID8
$\Phi^\dagger(\mathbf{3}, \mathbf{1}, 7/2)$	FIELD, SPIN, 0, SU3, 3, U1, 7, /, 2, DAGGER, ID2
	CONTRACTIONS
$\epsilon^{a_1 b_2 c_2}\ \phi_{a_1}\ \phi^\dagger_{b_1 b_2}\ \Phi_{c_1 c_2}\ \Phi^\dagger_{d_1}$	SU3, ID5, ID7, ID8
$\epsilon^{b_1 c_1 d_1}\ \phi_{a_1}\ \phi^\dagger_{b_1 b_2}\ \Phi_{c_1 c_2}\ \Phi^\dagger_{d_1}$	SU3, ID7, ID8, ID2
$\epsilon^{a_1 b_1}\ \phi_{a_1 a_2}\ \phi^\dagger_{b_1 b_2}\ \Phi\ \Phi^\dagger$	SU2, ID5, ID7
$\epsilon^{a_2 b_2}\ \phi_{a_1 a_2}\ \phi^\dagger_{b_1 b_2}\ \Phi\ \Phi^\dagger$	SU2, ID5, ID7

But why?

It's all about what comes next!

"What you did is a solved problem!"

BSM Extension Recommender

(Ongoing)

Known Fields & Extra Fields Requirement → BART-L+ → BSM Extension

Possible Directions

Lagrangian Autocomplete: Incomplete Lagrangian → AI → Complete Lagrangian

Symmetry Finder: Particle Content → AI → Lagrangian w/ Extra Symmetry

Symmetry Restoration: Symmetry Broken Lagrangian → AI → Symmetry Restored Lagrangian

Direct Data BSM Inference :
Large Language Model → pictures to words
Large Lagrangian Model → experimental data to Lagrangians

Experimental Data! → AI → Beyond Standard Model Lagrangian

Foundational model for *all* of particle physics : It must incorporate symbolic expressions - the language we use to describe and make sense of the universe – as a modality

Numerical Data ← Foundation models → Symbolic Language

Performance : How good is it?

Accuracy ≥90%

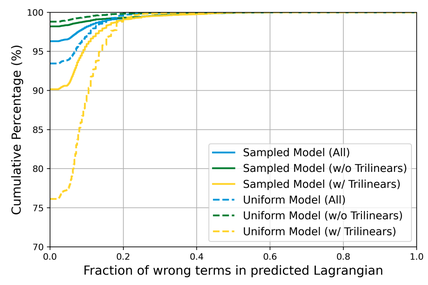
(Percentage of Dataset with $S_{Lagrangian} = 1$)

	Sampled Model	Uniform Model
Sampled Dataset	93.4%	89.9%
Uniform Dataset	90.5%	96.6%
Merged Dataset	92.0%	93.2%

$$S_{Lagrangian} = \frac{N^{Terms}_{Correct} - N^{Terms}_{Extra}}{N^{Terms}_{Expected}}$$

Tend to miss Yukawa Terms due to lack of Lagrangians involving many Yukawa Terms in training data (Solvable!)

Examples	Best $S_{Lagrangian}$
Lepton Sector	1
Quark Sector	1
Higgs + Leptons	1
Higgs + Quarks	0.86
Standard Model	0.77
2HDM	0.81
Georgi-Machacek	0.90
RPV-MSSM LLE	1
RPV-MSSM UDD	0.88
RPV-MSSM LQD	0.78

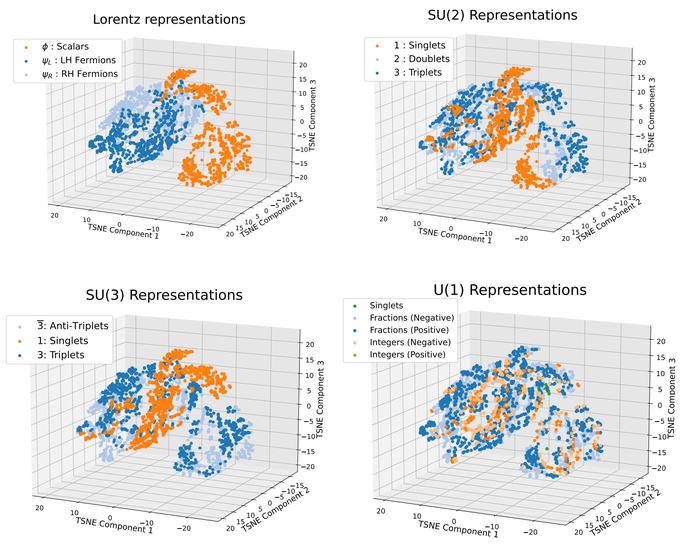


Embedding Analysis : What did it learn?

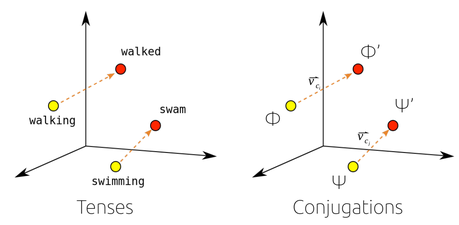
Symmetries!

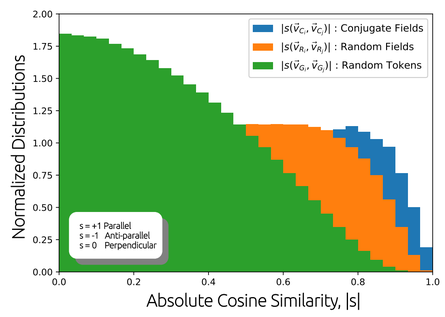
Despite being trained only to generate Lagrangians, the model has learned symmetry representations and conjugation transformations!

Field Embeddings



Vector Offset Method

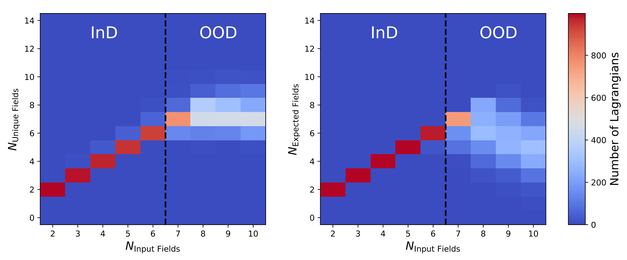




Out-of-Distribution (OOD) : Can it work with...

...more fields? YES! But it can't count properly

In-Distribution (Train) : ≤6 Fields
Out-of-Distribution (Test) : 7-10 Fields

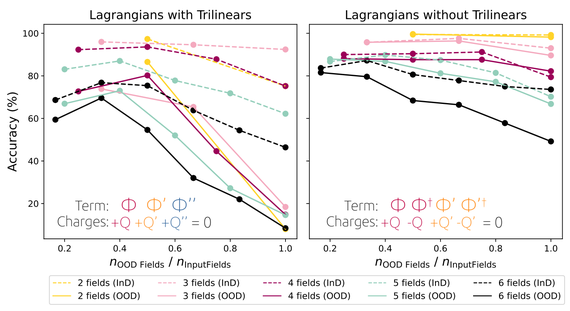


It knows there should be more fields in the Lagrangian, but it doesn't know how many (plateau at ~7)

BART's bidirectional encoder are known to struggle with contextual counting tasks [3]

...non-minimal charges? YES! But it can't do arithmetic

In-Distribution (Train) : Q = ½, 2, -7/2, ...
Out-of-Distribution (Test) : Q = 2/4, 6/2, 3/9, ...



It struggles with Lagrangians involving Trilinears because arithmetic is required for charge conservation in Trilinear terms while it's just pattern recognition for others!

Text-based encoding schemes for numbers are prone to take advantage of spurious correlations in the data [4]

Acknowledgement

The models were trained on Google Cloud Platform using research credits awarded to this project. The computations and data handling were enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022-06725. Specifically, we thank Chalmers e-Commons at Chalmers and the PDC Center for High Performance Computing at KTH Royal Institute of Technology, Sweden for providing access to the computing resources and data storage used in this research. We also thank the Lundström-Åman Scholarship for supporting travel to attend this conference.

References

1) Y.S. Koay, R. Enberg, S. Moretti and E. Camargo-Molina, Generating particle physics Lagrangians with transformers, , arXiv preprint arXiv:2501.09729 (2025).
2) R. V. Harlander and M. C. Schaaf, Autoeft: Automated operator construction for effective field theories, Computer Physics Communications 300, 109198 (2024).
3) S. Golkar, A. Bietti, M. Pettee, M. Eickenberg, M. Cranmer, K. Hirashima, G. Krawezik, N. Lourie, M. McCabe, R. Morel et al., Contextual counting: A mechanistic study of transformers on a quantitative task, arXiv preprint arXiv:2406.02585 (2024).
4) S. Golkar, M. Pettee, M. Eickenberg, A. Bietti, M. Cranmer, G. Krawezik, F. Lanusse, M. McCabe, R. Ohana, L. Parker, B. R.-S. Blancard, T. Tesileanu et al., xval: A continuous numerical tokenization for scientific language models (2024), 2310.02989.