Discovery and Application of Hidden Symmetries in Physics Data

- **Applicant:** Krutarth Parmar
- **Contact:** parmar.kay@icloud.com, https://github.com/kayo09, https://kayparmar.com/
- **Organization:** ML4SCI

- **Project Title:** Discovery of Hidden Symmetries and Conservation Laws using Physics-Aware Machine Learning
- **Mentors:**
    - Diptarko Choudhury (NISER)
    - Sergei Gleyzer (University of Alabama)
    - Ruchi Chudasama (University of Alabama)
    - Samuel Campbell (University of Alabama)
    - Emanuele Usai (University of Alabama)
    - Alex Roman (University of Alabama)

Abstract:

This project addresses the challenge of discovering hidden symmetries in physics datasets using machine learning, a crucial step towards building more robust, data-efficient, and interpretable physics-aware models. Symmetries fundamental to physics (like those in the Standard Model) can become obscured in complex data representations. This project will develop and benchmark ML techniques, inspired by recent literature **[1, 2, 3, 4]**, to first uncover rotational symmetry in a controlled MNIST

environment (VAE and MLP analysis) and then extend these methods to probe symmetries within high-energy physics contexts, potentially utilizing CMS open data. The project aims to implement unsupervised discovery algorithms, evaluate their effectiveness, and explore the use of discovered symmetries in constructing physics-aware models, aligning with the goals of advancing ML applications in scientific domains.

Motivation & Background:

Physics is deeply intertwined with the study of symmetry, linking invariance principles to fundamental conservation laws. While symmetries are often clear in standard formalisms (like 4-vectors), they can become "hidden" in complex, high-dimensional data representations common in experimental physics or abstract feature spaces learned by ML models. Automating the discovery of these hidden symmetries is a key challenge and opportunity at the intersection of physics and machine learning.

Recent progress in unsupervised and physics-aware ML provides powerful tools for this task. Techniques aiming to learn Lie group structures, minimize symmetry-violating PDE terms, or use adversarial methods offer paths to identify symmetries directly from data. Successfully learning these symmetries enables the construction of models that inherently respect physical constraints, leading to improved performance and generalization.

This project directly tackles this challenge. My motivation is driven by a strong interest in applying advanced ML to fundamental science problems. Having already completed the initial VAE implementation (Task 1) and the supervised MLP-based latent space analysis (Task 2), I am particularly keen to delve into the unsupervised techniques described in the reference papers and understand how symmetries can be discovered without prior labels. The prospect of potentially applying these methods to real-world physics data, like CMS datasets, and contributing to the development of truly physics-aware AI systems is particularly exciting. This GSoC project is an excellent platform to deepen my expertise in both ML and its scientific applications.

Project Goals & Deliverables:

1. **Task 1: Foundational Latent Space (Completed)**
    - **Goal:** Prepare rotated MNIST dataset (digits 1&2) and train a VAE baseline to create a latent space embedding rotational variations.
    - **Status:** Completed, including implementation, training, and initial analysis.
    - **Deliverable:** Documented code, trained VAE model, baseline analysis.
2. **Task 2: Supervised Symmetry Characterization**
    - **Goal:** Implement and train an MLP on the VAE's latent space to explicitly learn the transformation corresponding to known 30-degree image rotations.
    - **Status:** Implementation and initial training completed. GSoC time will involve thorough analysis, documentation, and potential refinement.
    - **Deliverable:** Documented MLP implementation and training code; trained MLP model; analysis report on the learned latent space transformation.
3. **Task 3: Unsupervised Symmetry Discovery (Next Step)**
    - **Goal:** Explore and implement unsupervised techniques, drawing inspiration from the reference papers (e.g., focusing on learning Lie algebra generators, using GANs, or PDE minimization), to discover rotational symmetry in the MNIST latent space without explicit angle labels, possibly by enforcing invariance of a property like classification logits.
    - **Deliverable:** Documented implementation of at least one unsupervised discovery algorithm based on the literature;

code; analysis comparing the discovered symmetries to the known rotation symmetry.

4. **Task 4: Extension to Physics Use Case (e.g., CMS Data)**
    - **Goal:** Adapt and apply the successfully implemented unsupervised discovery technique(s) from Task 3 to a more complex physics dataset. This could involve using simplified CMS open data or specific physics toy datasets to probe for known symmetries (e.g., Lorentz, phase space symmetries).
    - **Deliverable:** Code adapted for the chosen physics dataset format; analysis report detailing symmetries investigated and found.
5. **Task 5: Physics-Aware Model & Benchmarking**
    - **Goal:** Utilize the discovered symmetries (from Task 3/4) to inform the design of a simple physics-aware model (e.g., an invariant classifier or regressor). Benchmark this model against baselines in terms of data efficiency and/or invariance properties.
    - **Deliverable:** Prototype physics-aware model code; benchmarking results and analysis report.

Timeline (350h Project):

- **Community Bonding Period:** Deepen study of reference papers for Task 3; finalize documentation for Tasks 1 & 2; refine detailed plan for Tasks 3, 4 & 5 with mentors.
- **Weeks 1-6:** Task 3 - Implement and test first unsupervised discovery method on MNIST latent space.
- **Weeks 7-10:** Task 3 - Refine Task 3 implementation, potentially explore second method; complete analysis for Task 3.
- **Weeks 11-15:** Task 4 - Adapt discovery method(s) for physics dataset; perform analysis on physics data.

- **Weeks 16-20:** Task 5 - Design and implement physics-aware model; run benchmarking experiments and finalize analysis.
- **Final Weeks:** Complete code documentation and reports; prepare final GSoC submission.

About Me / Qualifications:

- I am a tech enthusiast
- I have proficiency in **Python** and experience with **PyTorch**, demonstrated through the completion of Task 1 (VAE implementation) and Task 2 (MLP on latent space). I am also proficient in **C++** and familiar with tensorflow and other machine learning frameworks.
- My previous experience in Deep Learning includes working on a functional approach to Pytorch classifier by implementing it using tinygrad framework, culminating in the successful implementation of the first two tasks of this project proposal.
- My background includes post graduate experience in machine learning and artificial intelligence from Caltech CTME
- I am highly motivated to learn the advanced unsupervised ML techniques central to Task 3 (Lie Group/Algebra representations, GANs, physics-informed losses) and apply them to challenging scientific problems, building upon the foundation laid by completing Tasks 1 and 2. I am committed to thoroughly studying the provided reference papers.

Commitment:

I am prepared to dedicate approximately 350 hours to this project during the GSoC 2025 period. I am enthusiastic about collaborating with mentors, actively participating in the ML4SCI community, and delivering high-quality results, continuing the progress I have already made.

Having completed the initial VAE and supervised MLP stages (Tasks 1 & 2), I am excited to proceed with the unsupervised

discovery phase (Task 3) and contribute further to this meaningful project with ML4SCI.

References:

1. Z. Liu and M. Tegmark, "Machine-learning hidden symmetries," arXiv:2109.09721v2 [cs.LG], May 2022.
2. R. Forestano, K. T. Matchev, K. Matcheva, A. Roman, E. Unlu, and S. Verner, "Deep Learning Symmetries and Their Lie Groups, Algebras, and Subalgebras from First Principles," arXiv:2301.05638v1 [hep-ph], Jan 2023.
3. J. Yang, R. Walters, N. Dehmamy, and R. Yu, "Generative Adversarial Symmetry Discovery," arXiv:2302.00236v4 [cs.LG], Jun 2023.
4. Roman, Alexander, et al. “Oracle-Preserving Latent Flows.” *Symmetry*, vol. 15, no. 7, July 2023, p. 1352. https://doi.org/10.3390/sym15071352.