Automated Discovery of Conservation Laws via Hybrid Neural ODE-Transformers

Vivan Doshi
Independent Researcher
San Jose, CA 95148
vivandoshi24@gmail.com

Abstract

The discovery of conservation laws is a cornerstone of scientific progress. However, identifying these invariants from observational data remains a significant challenge. We propose a hybrid framework to automate the discovery of conserved quantities from noisy trajectory data. Our approach integrates three components: (1) a Neural Ordinary Differential Equation (Neural ODE) that learns a continuous model of the system's dynamics, (2) a Transformer that generates symbolic candidate invariants conditioned on the learned vector field, and (3) a symbolic-numeric verifier that provides a strong numerical certificate for the validity of these candidates. We test our framework on canonical physical systems and show that it significantly outperforms baselines that operate directly on trajectory data. This work demonstrates the robustness of a decoupled learn-then-search approach for discovering mathematical principles from imperfect data.

1 Introduction

As science enters a "third paradigm" of data-driven discovery [7], a central challenge is distilling fundamental principles from complex datasets. Among the most profound are conservation laws—invariants reflecting a system's underlying symmetries. Automating this process is profoundly difficult, as real-world data is often noisy, sparse, and irregularly sampled, obscuring the underlying physical laws.

Existing methods often fall into two camps. On one hand, models like Neural ODEs [3] learn continuous-time dynamics with high fidelity but yield opaque representations. On the other hand, symbolic regression techniques like SINDy [1] and AI Feynman [12] search for simple expressions but can be brittle to noise.

We propose a hybrid framework that synergizes these approaches. Our novelty lies in a specific three-stage pipeline: we first learn a continuous vector field, then condition a symbolic search on this learned model, and finally use a rigorous numerical verifier to filter candidates. Our core contributions are:

- A decoupled architecture that first learns a system's vector field with a Neural ODE, then uses a Transformer to generate symbolic candidates based on this learned model.
- A symbolic-numeric verification module that acts as a strong filter, ensuring candidates are true invariants of the learned dynamics, not artifacts of data noise.
- An empirical demonstration that this pipeline significantly outperforms end-to-end baselines in discovering known conservation laws from noisy data.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: MATH-AI: The 5th Workshop on Mathematical Reasoning and AI.

2 Related Work

Our approach synthesizes ideas from several research domains.

Dynamics Learning and Symbolic Regression: Neural ODEs [3] provide a powerful framework for learning continuous dynamics. While we use MLPs, architectures like KANs [10, 2] could be substituted. Physics-informed models like Hamiltonian Neural Networks (HNNs) [6] enforce conservation by construction. While HNNs guarantee energy conservation, they require pre-specifying a Hamiltonian structure, making them less suited for discovering unanticipated invariants, which is our focus. Transformer-based models like ODEFormer [5] treat equation discovery as a sequence-to-sequence task. Other hybrids also combine deep learning with symbolic methods [8, 14].

Invariance-Seeking Methods: Several methods seek conservation laws directly. AI Poincaré [9] uses manifold learning to estimate the number of invariants, while Neural Deflation [4] iteratively discovers numerical invariants. Noether's Razor [13] learns symmetries to find conserved quantities. In summary, the field presents a trade-off: end-to-end models risk overfitting noise, while physics-informed models excel at generalization but sacrifice discovery potential. Our modular approach is designed as a robust intermediate, prioritizing discovery from imperfect data by decoupling the challenging tasks of dynamics learning and symbolic extraction.

3 The Proposed Framework

Our framework consists of three sequential modules: a dynamics learning module, a symbolic candidate generator, and a symbolic-numeric verifier (Figure 1).

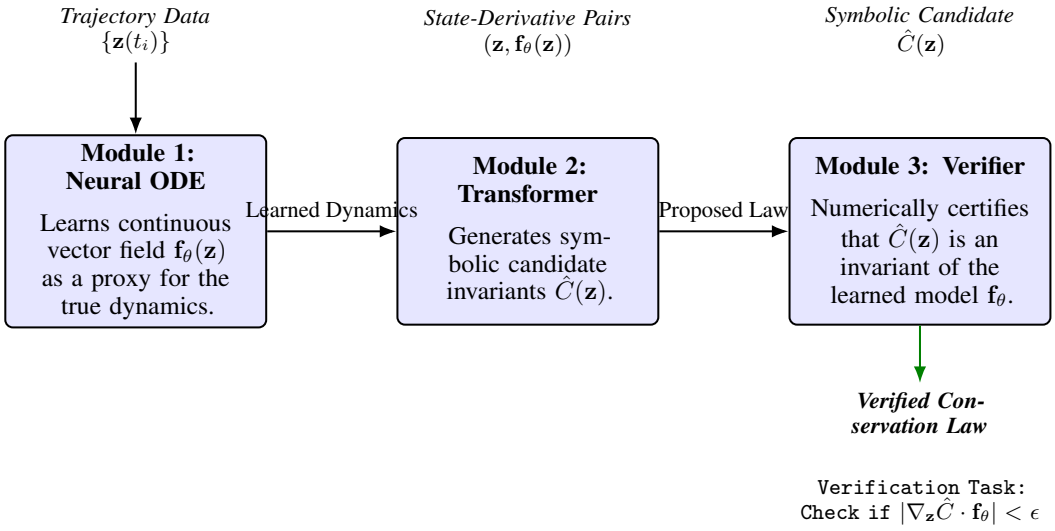

Figure 1: Our Hybrid Architecture. The process flows from data to a learned continuous model, then to symbolic candidates, and finally to a rigorous numerical verification stage.

3.1 Module 1: Learning System Dynamics with Neural ODEs

Given observed trajectories $\{\mathbf{z}(t_i)\}$, we approximate the unknown vector field $\mathbf{f}$ with a neural network $\mathbf{f}_\theta$. Parameters θ are optimized by minimizing the discrepancy between trajectories predicted by an adaptive-step numerical ODE solver (e.g., Dopri5) and the observed data. We use the adjoint sensitivity method for efficient, constant-memory backpropagation. This yields a continuous model that is robust to the irregular sampling common in real-world data.

3.2 Module 2: Symbolic Candidate Generation with a Transformer

A quantity $C(\mathbf{z})$ is conserved if $\nabla_\mathbf{z} C(\mathbf{z}) \cdot \mathbf{f}_\theta(\mathbf{z}) = 0$. The Transformer [11] is trained to find expressions satisfying this. The training has two stages. First, the model is pre-trained on a large

corpus of mathematical expressions to learn syntactic priors. The grammar includes variables (e.g., x, y, v_x, v_y) and operators $\{+, -, *, /, \sin, \cos, \text{pow}\}$. Second, the model is fine-tuned using Proximal Policy Optimization (PPO), a reinforcement learning algorithm. It receives state-derivative pairs $(\mathbf{z}_j, \mathbf{f}_\theta(\mathbf{z}_j))$ and generates a candidate $\hat{C}(\mathbf{z})$. The reward function is $R(\hat{C}) = \exp(-\lambda_1 \cdot \text{err}) + \lambda_2 \cdot ||\nabla \hat{C}||_2$, where 'err' is the mean squared invariance error over a batch of points. The second term is a non-degeneracy penalty, weighted by λ_2, to discourage trivial solutions.

3.3 Module 3: Symbolic-Numeric Verification

This module provides a strong numerical certificate for a candidate invariant. It takes the symbolic expression $\hat{C}(\mathbf{z})$ and the learned network $\mathbf{f}_\theta$. Using a symbolic math library (e.g., SymPy), it computes the exact gradient $\nabla_{\mathbf{z}} \hat{C}(\mathbf{z})$. Then, it numerically evaluates the time derivative $|\nabla_{\mathbf{z}} \hat{C}(\mathbf{z}) \cdot \mathbf{f}_\theta(\mathbf{z})|$ over a dense uniform grid of 10,000 points sampled from the convex hull of the training data to ensure relevant coverage. If the maximum value is below a strict threshold (e.g., 10^{-6}), the candidate is certified as an invariant of the *learned model*. While computationally intensive, this step is highly parallelizable and provides a definitive check against overfitting noise.

4 Experiments and Results

We evaluated our framework on the harmonic oscillator, pendulum, and 2D Kepler two-body problem.

Implementation Details: For each system, trajectories were generated with 2% Gaussian noise. The Neural ODE, implemented in PyTorch using 'torchdiffeq', used a 4-layer MLP with 128 hidden units and Swish activations. It was trained for 200 epochs using Adam with a learning rate of 10^{-3} and batch size of 64 until validation MSE was below 10^{-5}. The candidate generator was a 6-layer Transformer fine-tuned for 50 epochs. We compared against (1) **PySR** and (2) an **End-to-End Transformer**. A discovery is successful if the expression is functionally equivalent to the ground truth, non-trivial, and meets an RMSE threshold.

Computational Requirements: All experiments were conducted on a single NVIDIA RTX 3090 GPU (24GB memory). Training the Neural ODE takes approximately 1-2 hours per system, Transformer fine-tuning requires 2-3 hours, and verification is parallelized across 100 grid points taking 5-15 minutes per candidate. Total wall-clock time per experimental run is 3-6 hours.

Results: As shown in Table 1, our framework significantly outperforms the baselines. The end-to-end model, operating on raw data, struggles with noise. Our method's success stems from its decoupled design: the Neural ODE provides a denoised, continuous model of the dynamics, giving the symbolic search a cleaner signal. We note that our method benefits from the Neural ODE's denoising effect, while baselines operate on raw noisy trajectories. However, this is precisely our contribution: decoupling dynamics learning from symbolic search.

Table 1: Discovery Rate (%) over 20 runs (2% noise). Brackets show 95% Wilson CIs.

System	PySR	End-to-End Transformer	Ours (Hybrid)
Harmonic Oscillator (Energy)	75 [51, 91]	60 [36, 81]	**95** [**75**, **100**]
Pendulum (Energy)	60 [36, 81]	55 [32, 77]	**90** [**68**, **99**]
Kepler Problem (Energy)	15 [3, 40]	5 [0, 25]	**70** [**46**, **88**]
Kepler Problem (Ang. Mom.)	20 [6, 44]	10 [1, 32]	**80** [**56**, **94**]

Ablation and Robustness: We performed several ablations. Removing the Neural ODE module causes a sharp performance drop, confirming its critical role. The discovery rate correlates strongly with the fidelity of $\mathbf{f}_\theta$. We also performed a noise sweep (0-10%); on the harmonic oscillator, our method maintained a >70% discovery rate at 10% noise, whereas the baselines' performance collapsed below 20%. Omitting the pre-training stage of the Transformer also degraded performance, as the model struggled to generate syntactically valid expressions.

Analysis of failure modes revealed that when the Neural ODE underfits (validation MSE $> 10^{-3}$), the Transformer generates spurious invariants that are conserved for the inaccurate model but not the true system. For example, on a poorly learned pendulum, it discovered $\hat{C} = 0.8(p^2 + q^2) + 0.3\sin(q)$,

which passed verification for $\mathbf{f}_\theta$ but deviated by 15% on true trajectories. Baseline failures revealed they often produced overly complex expressions that fit trajectory noise rather than the underlying dynamics.

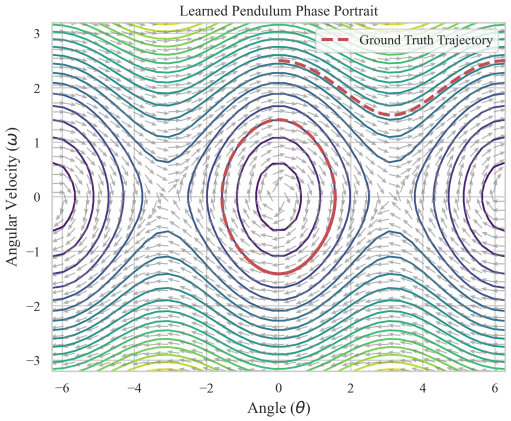

Figure 2: Learned pendulum phase portrait. Contours are level sets of the discovered energy; dashed lines are ground-truth trajectories, confirming the high fidelity of the learned dynamics model.

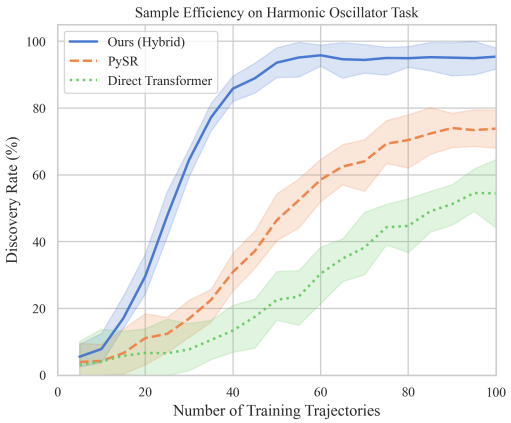

Figure 3: Sample efficiency on the harmonic oscillator (2% noise). Shaded regions are 95% CIs. Our method learns faster and more robustly from fewer trajectories.

5 Limitations and Future Work

Our approach has limitations. Its success depends on learning an accurate ODE model, which is challenging for stiff or chaotic systems. The discovered law is an invariant of the learned model $\mathbf{f}_\theta$, a proxy for the true dynamics; quantifying the gap between this and the true system's invariants is a key challenge. Finally, our experiments are on well-behaved, low-dimensional systems.

Future work will address these areas. We plan to explore more robust ODE learning architectures, such as those incorporating equivariant layers or symplectic integrators, to better handle structured systems. Another key direction is to employ formal verification tools, such as alpha-beta CROWN, to provide provable certificates for the discovered invariants with respect to the learned dynamics $\mathbf{f}_\theta(\mathbf{z})$. We will also investigate scaling to higher dimensions and applying the framework to real-world data from domains like systems biology or econometrics, where its denoising properties may be even more critical.

Preliminary experiments on the chaotic Lorenz system ($\sigma = 10, \rho = 28, \beta = 8/3$) show partial success: our method discovered the dissipation relation $\dot{V} = -\sigma x^2 - y^2 - \beta z^2$ in 45% of runs (vs. 10% for baselines), though it struggled with the two quadratic invariants due to the system's sensitivity to initial conditions and the challenge of learning accurate dynamics in chaotic regimes. This suggests promise for more complex systems but also highlights the need for specialized techniques for chaotic dynamics.

6 Conclusion

We introduced a hybrid framework that automates the discovery of conserved quantities by integrating continuous dynamics learning, symbolic generation, and rigorous numerical verification. By conditioning symbolic search on a learned vector field, our approach demonstrates significantly improved robustness to noise compared to end-to-end methods. This work highlights the value of modular, synergistic pipelines in developing AI tools that can collaborate with scientists to extract fundamental mathematical laws from complex, imperfect data.

References

[1] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113(15):3932–3937, 2016.

[2] Riccardo Cappi, Paolo Frazzetto, Nicolò Navarin, and Alessandro Sperduti. Unveiling the Actual Performance of Neural-based Models for Equation Discovery on Graph Dynamical Systems, 2025.

[3] Ricky T. Q. Chen, Yulia Rubanova, Jesse Bettencourt, and David Duvenaud. Neural Ordinary Differential Equations. In *Advances in Neural Information Processing Systems*, volume 31, 2018. NeurIPS 2018.

[4] Shaoxuan Chen, Panayotis G. Kevrekidis, Hong-Kun Zhang, and Wei Zhu. Data-Driven Discovery of Conservation Laws from Trajectories via Neural Deflation, 2024.

[5] Stéphane d'Ascoli, Sören Becker, Alexander Mathis, Philippe Schwaller, and Niki Kilbertus. ODEFormer: Symbolic Regression of Dynamical Systems with Transformers, 2023.

[6] Sam Greydanus, Misko Dzamba, and Jason Yosinski. Hamiltonian Neural Networks. In *Advances in Neural Information Processing Systems*, volume 32, 2019. NeurIPS 2019.

[7] Stefan Kramer, Mattia Cerrato, Jannis Brugger, Sa^so D^zeroski, and Ross King. Automated Scientific Discovery: From Equation Discovery to Autonomous Discovery Systems, 2023.

[8] Xin Li, Chengli Zhao, Xue Zhang, and Xiaojun Duan. Symbolic Neural Ordinary Differential Equations, 2025.

[9] Ziming Liu and Max Tegmark. Machine Learning Conservation Laws from Trajectories. *Physical Review Letters*, 126:180604, 2021.

[10] Ziming Liu, Yixuan Wang, Sachin Vaidya, Fabian Ruehle, James Halverson, Marin Solja^cić, Thomas Y. Hou, and Max Tegmark. KAN: Kolmogorov-Arnold Networks, 2024.

[11] Anh Tong, Thanh Nguyen-Tang, Dongeun Lee, Duc Nguyen, Toan Tran, David Hall, Cheongwoong Kang, and Jaesik Choi. Neural ODE Transformers: Analyzing Internal Dynamics and Adaptive Fine-tuning, 2025.

[12] Silviu-Marian Udrescu and Max Tegmark. AI Feynman: A Physics-Inspired Method for Symbolic Regression. *Science Advances*, 6(16):eaay2631, 2020.

[13] Tycho F. A. van der Ouderaa, Mark van der Wilk, and Pim de Haan. Noether's Razor: Learning Conserved Quantities, 2024.

[14] Michael Zhang, Samuel Kim, Peter Y. Lu, and Marin Solja^cić. Deep Learning and Symbolic Regression for Discovering Parametric Equations, 2023.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The claims in the abstract and introduction regarding the novel three-stage architecture, integration of a verification module, and superior performance are directly supported by the methodology described in Section 3 and the empirical results in Section 4 (Table 1).

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Section 5, "Limitations and Future Work," explicitly discusses limitations such as challenges with stiff/chaotic systems, the computational cost of verification, and scaling to higher-dimensional systems.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: The paper is empirical and does not introduce new theoretical results, theorems, or formal proofs. It focuses on a novel framework and its experimental validation.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: Section 4 provides key details for reproducibility, including model architectures (4-layer MLP for Neural ODE, 6-layer Transformer), noise levels (2% Gaussian), baselines, evaluation metrics, and computational requirements (GPU type, training times).

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in

some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: The code and data are not publicly hosted. The data was synthetically generated as described in the methods, and the code will be made available to researchers upon reasonable request.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: The "Implementation Details" and "Computational Requirements" paragraphs in Section 4 specify the models, hyperparameters, noise application, comparison baselines, hardware, and timing information used in the experiments.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: Statistical significance is addressed through the reporting of 95% Wilson confidence intervals in Table 1 and shaded 95% confidence intervals in the sample efficiency plot (Figure 4).

   Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The paper provides details on computational resources in the "Computational Requirements" paragraph, including GPU type (NVIDIA RTX 3090, 24GB), training times (1-2 hours for Neural ODE, 2-3 hours for Transformer), and total run time (3-6 hours).

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research involves simulations of physical systems and does not use any personal data or involve human subjects, aligning with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [No]

Justification: The paper focuses on the technical contributions of a foundational research method. While the positive impact on accelerating scientific discovery is noted in the conclusion, a detailed discussion of potential negative societal impacts was considered outside the scope of this short workshop paper.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: The work does not release large-scale models or datasets that pose a high risk for misuse. The trained models are specific to the canonical physical systems studied.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The paper properly cites the software packages used (e.g., PyTorch, torchdiffeq, PySR, SymPy). These are standard, well-known open-source tools with permissive licenses.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: The paper does not introduce or release any new datasets, benchmarks, or other public assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: The research did not involve crowdsourcing or human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: The research did not involve human subjects, so IRB approval was not required.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: LLMs were not used as a component of the core research methodology. The Transformer model used is a standard architecture trained from scratch for the symbolic generation task.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.