A Personalised Learning Tool for Physics Undergraduate Students Built On a Large Language Model for Symbolic Regression

1st Yufan Zhu
School of Computing
National University of Singapore
Singapore, Singapore
e0773591@u.nus.edu

2nd Zi-Yu Khoo
School of Computing
National University of Singapore
Singapore, Singapore
e0395550@u.nus.edu

3rd Jonathan Sze Choong Low
Agency for Science, Technology and Research (A*STAR)
Singapore Institute of Manufacturing Technology
Singapore, Singapore
sclow@simtech.a-star.edu.sg

4th Stéphane Bressan
School of Computing
National University of Singapore
Singapore, Singapore
steph@nus.edu.sg

Abstract—Interleaved practice enhances the memory and problem-solving ability of students in undergraduate courses. We introduce a personalized learning tool built on a Large Language Model (LLM) that can provide immediate and personalized attention to students as they complete homework containing problems interleaved from undergraduate physics courses. Our tool leverages the dimensional analysis method, enhancing students' qualitative thinking and problem-solving skills for complex phenomena. Our approach combines LLMs for symbolic regression with dimensional analysis via prompt engineering and offers students a unique perspective to comprehend relationships between physics variables. This fosters a broader and more versatile understanding of physics and mathematical principles and complements a conventional undergraduate physics education that relies on interpreting and applying established equations within specific contexts. We test our personalized learning tool on the equations from Feynman's lectures on physics. Our tool can correctly identify relationships between physics variables for most equations, underscoring its value as a complementary personalized learning tool for undergraduate physics students.

Index Terms—AI and Education, Symbolic Regression, Large Language Models, Physics Education, Prompt Engineering, Undergraduate Learning

I. INTRODUCTION

Interleaved practice enhances students' memory and problem-solving ability in undergraduate physics courses [1]. This involves students alternating between topics in physics while completing their homework assignments. However, different students learn at different paces, and tutors cannot simultaneously provide complete attention to different students as they work on different topics. The need for a personalized learning tool for each student can be met with Large Language Models (LLMs).

Dimensional analysis, a fundamental concept in physics education, is renowned for its efficacy in cultivating qualitative thinking and unraveling complexities in physics. Its robust application in higher education, as demonstrated in advanced topics like Rayleigh Scattering, underscores its versatility and power. While foundational works by pioneers like Buckingham and Bridgman have cemented its importance, recent studies suggest the potential for its enhanced application in teaching methodologies [2]–[6]. The integration of dimensional analysis into an LLM-driven learning environment further leverages this powerful tool, deepening students' understanding of physics beyond conventional pedagogical boundaries.

LLMs can facilitate the learning of physics from data and the inferencing of laws by leveraging symbolic regression to distill complex physics laws from raw data [7]. This method aligns well with the historical context of physics discovery, where major theories often emerged from extensive data analysis [8], [9]. Incorporating these algorithms into LLMs allows for a more data-driven and exploratory approach to learning physics, enabling students to infer and understand the underlying principles from empirical data.

In this context, our research introduces a personalized learning tool built on an LLM that leverages dimensional analysis and symbolic regression. This tool is designed to cater to the individual learning pace of students, guiding them through interleaved practice in undergraduate physics courses. Our approach aims to enhance the conventional physics education model, which often limits students to memorizing and applying established equations within specific contexts.

We experimentally evaluate our personalized learning tool by applying it to equations drawn from Feynman's lectures on physics [10]–[12]. Our tool possesses a remarkable capacity to discover and establish relationships between various physics variables accurately. This efficacy was observed across the majority of the equations tested, which underscores the tool's potential as a beneficial resource in undergraduate physics education. Our tool offers a customized learning experience,

adapting to individual student needs and learning styles. Consequently, this tool stands as a valuable complement to traditional teaching methods, enhancing students' grasp of complex physics concepts and their application in various scenarios.

II. BACKGROUND

A. Symbolic Regression

Symbolic regression is a process that seeks to discover an explicit symbolic formula, representing a mapping from a given set of numerical input-output pairs $\{x_i, y_i\}_{i=1}^{M}$, where each x_i is a vector in a d-dimensional space ($x_i \in \mathbb{R}^d$) and each y_i is a corresponding output in a q-dimensional space ($y_i \in \mathbb{R}^q$). The core objective is to find a symbolic expression, denoted as f, that accurately describes the mathematical relationship $y = f(x)$ inherent in the data [13]. Unlike traditional regression models that might yield complex numerical algorithms, symbolic regression strives to provide a clear, human-readable formula that captures the intricacies and underlying patterns of the dataset.

Symbolic regression is an NP-hard problem, requiring a complex search through an extensive space of mathematical expressions to find the one that best fits a dataset. This complexity involves not just identifying the appropriate mathematical operations but also their optimal arrangement and combination. The resultant expression, f, provides a model that reveals insights into the underlying relationship between variables [14].

B. Dimensional Analysis

Dimensional analysis is a powerful tool in physics. It is used to simplify complex problems by focusing on the dimensions of physical quantities. The dimensions of a physical quantity Q are expressed as $[Q] = M^a L^b T^c$, where $[Q]$ denotes the dimensions of Q in terms of mass (M), length (L), and time (T). The utility of dimensional analysis extends beyond mere simplification. Historical contributions like Buckingham's study on physically similar systems and Bridgman's seminal book on dimensional analysis establish the method's foundational role in physics research and education [2].

The importance of dimensional analysis in education is further emphasized by its ability to develop students' intuitive understanding of physical phenomena. Furthermore, Blasiak et al. and Taber have also discussed the multi-stage nature of physics comprehension and the importance of sequencing in education, suggesting a need for innovative methods that integrate dimensional analysis more effectively into physics learning [6], [15]. Dimensional analysis has thus emerged as a crucial tool in developing a deeper and more intuitive understanding of physics, underscoring the need for its effective integration into physics curricula.

Rayleigh's method is an insightful technique within the dimensional analysis literature that seeks to elucidate the relationships between various physical quantities [16]. It begins by identifying all independent variables that may influence a dependent variable. The method then formulates a functional relationship between these variables, generally as a power product, taking the form:

$$R = C R_1^a R_2^b R_3^c \ldots R_n^m, \tag{1}$$

where R is the dependent variable, $R_1, R_2, \ldots, R_n$ are the independent variables, C is a dimensionless constant, and $a, b, c, \ldots, m$ are the exponents to be determined.

Each variable is expressed in base units, and through the principle of dimensional homogeneity, a set of simultaneous equations is formulated. Solving these equations yields the values of the exponents and forms a dimensionally consistent equation relating the independent and dependent variables.

C. Large Language Models in Physics Education

LLMs like the Generative Pre-trained Transformer (GPT) have shown significant potential to revolutionize education, especially in fields like physics. These models can analyze and generate text-based content, aiding in problem-solving and explanation generation [17]. In physics education, LLMs have been utilized for solving calculation problems, explaining concepts, and creating new exercises, offering a novel approach to learning and understanding complex topics [18]. However, the effective use of LLMs in physics education requires careful consideration of their capabilities and limitations. Recent studies have highlighted both the potential and the challenges of integrating LLMs into physics education, particularly in terms of accuracy, reliability, and contextual understanding [19].

III. RELATED WORKS

A. Symbolic Regression

Symbolic regression aims to find mathematical models fitting empirical data, evolving significantly since its early days [20]. Current algorithms are mainly heuristic or exact, with heuristic approaches prioritizing speed and exact methods focusing on accuracy [21].

Conventional symbolic regression uses genetic programming, an evolutionary technique derived from genetic algorithms and inspired by natural selection [22]. Pioneered by Holland [23], and extended by Cramer [24], this method involves evolving populations of candidate programs, modified through genetic operations like crossover and mutation. Koza's work further demonstrated the effectiveness of genetic algorithms in diverse problem domains, including symbolic regression [25]

Neural networks have revolutionized symbolic regression, leveraging their strength in gradient-based optimization and handling high-dimensional data [26], [27] to enable the discovery of intricate parametric equations [28], [29].

Physics-informed symbolic regression integrates physical principles to enhance expression discovery. A key development in this domain is AI Feynman, which employs neural networks to streamline equation discovery, effectively re-discovering equations from the Feynman Lectures on Physics and serving as a benchmark in the field [30]. Another notable advancement is the Scientist-Machine Equation Detector (SciMED) by Keren et al., an open-source framework that melds scientific

knowledge with advanced symbolic regression techniques for identifying physically meaningful symbolic expressions [31].

B. Prompt Engineering for Reasoning

Reasoning is a structured cognitive process that can be leveraged by LLMs [32] to augment their performance in symbolic regression. Effective prompt engineering, which tailors input prompts to guide LLM behavior, is key to maximizing their potential in reasoning tasks [33]. This involves single-stage and multi-stage approaches, discussed in Section III-B1 and III-B2 respectively.

1) Single-stage Approach: The single-stage approach, inspired by the proficiency of LLMs as few-shot reasoners, predominantly employs template-based prompts. Techniques like Contrastive Explanations and POTTER have harnessed prompts to augment commonsense reasoning with pre-trained models [34]. Notably, the Chain-of-Thought prompting approach introduced by Wei et al. uses a sequence of intermediate reasoning steps within few-shot prompts, making it especially beneficial for multi-step problems [35]. Additionally, other techniques have emerged that optimize prompt permutations, examine the impact of exemplar diversity, and incorporate explicit explanations to refine LLM performance in various tasks [36]. However, the single-stage approach efficacy can be sensitive to the selection of exemplars [37].

2) Multi-stage Approach: The multi-stage approach in prompt engineering, unlike its single-stage counterpart, employs a series of input-output exchanges to better guide LLMs in complex reasoning tasks. This method was substantiated by the foundational works of Kazemi et al. and Creswell and Shanahan, who conceptualized reasoning into sub-modules and developed a selection inference framework, respectively [38], [39]. Press et al. advanced this notion by integrating follow-up questions and intermediate answers, thus bridging the compositionality gap in LLMs [40]

In recent years, the multi-stage approaches have markedly shaped research, giving rise to various state-of-the-art techniques. This includes Generated-knowledge Prompting, which enhances LLMs by incorporating external knowledge for improved commonsense reasoning [41]. Alongside, Least-to-Most Prompting sequentially dissects complex problems, aiding in tasks that require progressive reasoning [42]. Adding to the repertoire, Self-Refine Prompting iteratively refines outputs through feedback loops, thus enhancing performance across various tasks [43].Each of these techniques represents a significant leap in the ongoing evolution of multi-stage approaches in LLMs.

IV. METHODOLOGY

Our methodology uses custom prompts to guide an LLM in generating dimensionally consistent physics equations. We empirically compare different methodologies that use single-stage or multiple-stage prompting approaches and evaluate their efficacy in guiding the model's output. Additionally, we have integrated established theorems, such as Rayleigh's method, to enrich the model's ability to generate physically

meaningful expressions, thereby enhancing its utility in both educational and research contexts.

A. Single-stage General Dimensional Analysis Method

This method utilizes a single-stage prompting approach for symbolic regression with an LLM. The LLM is presented with both regressors and response variables along with their respective dimensions, and a dataset containing variable values. In this single stage, the LLM is instructed to simultaneously propose dimensionally consistent equations through dimensional analysis and determine the coefficients of the proposed equation by fitting the data.

> **Sample Prompt**
> **Regressors:**
> $X1 : [I][T]$
> $X2 : [M][L][T]^-3[I]^-1$
> **Response Variable:**
> $Y : [M][L][T]^-2$
> Propose a dimensionally consistent equation form for the response variable as a function of the given regressors with unknown coefficients.

X1	X2	Y
4.230094105945206	4.989988020338535	21.10811891357122
1.7835543769151454	4.879775762873631	8.703345420237708
	...	

> Based on the data set given, derive the unknown coefficients for the equation form proposed in the previous step.

B. Two-stage General Dimensional Analysis Method

This method employs a two-stage approach to prompt symbolic regression. In Stage 1, the LLM is presented with both regressors and response variables, including their respective dimensions. The primary objective is to generate dimensionally consistent equations through dimensional analysis. Following the generation of the equation form, Stage 2 involves supplying LLMs with a dataset containing variable values to determine coefficients.

> **Sample Prompt for Stage One**
> **Regressors:**
> $X1 : [I][T]$
> $X2 : [M][L][T]^-3[I]^-1$
> **Response Variable:**
> $Y : [M][L][T]^-2$
> Propose a dimensionally consistent equation form for the response variable as a function of the given regressors with unknown coefficients.
> **Sample Prompt for Stage Two**
> **Data Points Table:**

X1	X2	Y
4.230094105945206	4.989988020338535	21.10811891357122
1.7835543769151454	4.879775762873631	8.703345420237708
	...	

> Based on the data set given, derive the unknown

coefficients for the equation form proposed in the previous step.

C. Single-stage Rayleigh's Method Analysis Method

This method utilizes a single-stage prompting approach, with the integration of Rayleigh's Method for Dimensional Analysis 1. The LLM is presented with both regressors and response variables along with their respective dimensions, and a dataset containing variable values. The LLM is instructed to simultaneously propose dimensionally consistent equations through Rayleigh's Method and determine the coefficients of the proposed equation by fitting the data.

Sample Prompt
Regressors:
$X1 : [I][T]$
$X2 : [M][L][T]^-3[I]^-1$
Response Variable:
$Y : [M][L][T]^-2$
Propose an equation $Y = C \cdot X_1^{a_1} \cdot X_2^{a_2} \cdot \ldots \cdot X_n^{a_n}$ using Rayleigh's Method, ensuring dimensional consistency. Solve for the unknown exponents $a_1, a_2, \ldots, a_n$ to make the equation dimensionally homogeneous.
Data Points Table:

X1	X2	Y
4.230094105945206	4.989988020338535	21.10811891357122
1.7835543769151454	4.879775762873631	8.703345420237708
	...	

Based on the data set given, derive the unknown coefficients for the equation form proposed in the previous step.

D. Two-stage Rayleigh's Method Analysis Method

This method employs a two-stage approach to prompt symbolic regression. In Stage 1, the LLM is presented with both regressors and response variables, including their respective dimensions. The primary objective is to generate dimensionally consistent equations through Rayleigh's Method. Following the generation of the equation form, Stage 2 involves supplying LLMs with a dataset containing variable values to determine coefficients.

Sample Prompt for Stage One
Regressors:
$X1 : [I][T]$
$X2 : [M][L][T]^-3[I]^-1$
Response Variable:
$Y : [M][L][T]^-2$
Propose an equation $Y = C \cdot X_1^{a_1} \cdot X_2^{a_2} \cdot \ldots \cdot X_n^{a_n}$ using Rayleigh's Method, ensuring dimensional consistency. Solve for the unknown exponents $a_1, a_2, \ldots, a_n$ to make the equation dimensionally homogeneous.
Sample Prompt for Stage Two
Data Points Table:

X1	X2	Y
4.230094105945206	4.989988020338535	21.10811891357122
1.7835543769151454	4.879775762873631	8.703345420237708
	...	

Based on the data set given, derive the unknown coefficients for the equation form proposed in the previous step.

V. PERFORMANCE EVALUATION

A. Experiment Setup and Evaluation Metrics

In our study, we utilized the GPT-4 model, serving as the LLM, and integrated it with a Wolfram Alpha Plugin. This combination significantly bolstered the model's computational abilities, particularly in performing mathematical and symbolic operations, essential for the scope of our research. To prevent the model from relying on prior physics knowledge, variables are obfuscated with generic names, e.g., $E = q_2 \times E_f$ becomes $Y = X_1 \times X_2$.

Our study utilizes 26 equations from the Feynman Lectures on Physics for undergraduate physics students [30].We focus on the equations delineated by Udrescu et al. that are solvable by dimensional analysis [30]. For all regression tasks, we used 20 data points per equation.

We specify the methodologies and metrics employed to assess the effectiveness and efficiency of the LLM.

The first metric comprises two counts which assesses the LLM's ability to suggest and regress equation forms efficiently and accurately. If the LLM fails to suggest the expected form, the same prompt is iteratively used until success. Therefore a smaller count is better. The first count is the number of prompts required for the LLM to suggest a dimensionally consistent equation. The second is the number of prompts needed to accurately regress the coefficients.

The second metric is the Mean Absolute Percentage Error (MAPE). It is an assessment of fit. The LLM regresses the coefficients, regardless of the dimensional consistency of the initially proposed model. The MAPE is averaged over all 26 equations in the study.

B. Performance Evaluation

We assess the four proposed methods utilizing the two aforementioned metrics. The findings, detailed in Table I, underscore the efficiency of LLMs in single-stage prompt resolutions. The success rate for the Single-stage General Dimensional Analysis Method underscores the robust capability of LLM in symbolic regression tasks even within the confines of a single-stage process. Furthermore, the Single-stage Rayleigh's Method Analysis Method requires fewer prompts on average than the Single-stage General Dimensional Analysis Method. The incorporation of an established dimensional analysis approach enhances the LLM's performance further. Lastly, the Two-Stage Rayleigh's Method Analysis Method had the highest success rate. The combination of an established dimensional analysis approach and a multi-stage prompting

	Methodology						
	Single-stage General Dimensional Analysis	Two-stage General Dimensional Analysis		Single-stage Rayleigh's Method Analysis	Two-stage Rayleigh's Method Analysis		AI Feynman
Successfully discovered in One Shot	17/26	22/26		17/26	23/26		26/26
Average No. of Prompts Needed to	Regress expected equation	Propose dimensionally consistent equation	Regress expected coefficients	Regress expected equation	Propose dimensionally consistent equations	Regress expected coefficients	N.A.
	1.885	1.154	1.192	1.308	1	1.192	N.A.
MAPE	105.86%	22.88%		9.95%	3.77%		$\approx 0\%$
Average Time Taken	2m33s	56s	2m42s	2m17s	1m6s	2m40s	18m50s

technique yields the best results. This refinement underscores the effectiveness of our methodology, particularly in tackling sophisticated symbolic regression tasks, demonstrating its practical utility in complex computational scenarios.

In comparison to the baseline symbolic regression AI Feynman, Table I shows that AI Feynman has a higher success rate but is very time-intensive. In contrast, our four methodologies consistently discovered equations within 5 minutes. This efficiency makes our tool more user-friendly and accessible, as students can receive prompt attention and guidance on physics problems, to match their learning paces.

VI. CONCLUSION

Our investigation into the application of LLMs for symbolic regression with dimensional analysis tasks in physics has yielded promising results. We observed that both the General Dimensional Analysis Approach and Rayleigh's Method Analysis Approach have proven effective, delivering equations with good accuracy. Notably, our study reveals that two-staged prompting approaches outperform single-stage prompting, underscoring the importance of structured problem-solving and systematic analysis in leveraging LLMs for educational purposes. Furthermore, the integration of established dimensional analysis methodology such as Rayleigh's Method into prompts has been shown to further enhance the effectiveness of symbolic regression. This synergy allows for a more refined exploration of complex physical phenomena, providing deeper insights that are invaluable in an educational context.

In conclusion, our research highlights the potential role of LLMs as innovative and accessible educational tools in undergraduate physics education. By harnessing the power of prompt engineering and combining it with scientific knowledge like Rayleigh's Method, LLMs facilitate personalized and data-driven learning experiences, allowing for a nuanced understanding of physical phenomena. Our approach not only deepens students' understanding of physics concepts but also nurtures essential critical thinking and analytical skills, pivotal in today's scientific arena. Consequently, LLMs emerge as valuable assets in the realm of undergraduate physics education, offering novel avenues for learning and exploration.

REFERENCES

[1] J. Samani and S. C. Pan, "Interleaved practice enhances memory and problem-solving ability in undergraduate physics," *npj Science of Learning*, vol. 6, no. 1, p. 32, Nov 2021. [Online]. Available: https://doi.org/10.1038/s41539-021-00110-x

[2] E. Buckingham, "On physically similar systems; illustrations of the use of dimensional equations," *Physical Review*, vol. 4, no. 4, pp. 345–376, 1914. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRev.4.345

[3] P. W. Bridgman, *Dimensional Analysis*. New Haven: Yale University Press, 1922. [Online]. Available: https://shorturl.at/cDM45

[4] L. Bao, K. Fang, T. Cai, J. Wang, L. Yang, L. Cui, J. Han, L. Ding, and Y. Luo, "Learning of content knowledge and development of scientific reasoning ability: A cross culture comparison," *American Journal of Physics*, vol. 77, no. 12, pp. 1118–1123, 2009. [Online]. Available: https://doi.org/10.1119/1.2976334

[5] M. Marušić and J. Sliško, "Effects of two different types of physics learning on the results of class test," *Phys. Rev. ST Phys. Educ. Res.*, vol. 8, p. 010107, Feb 2012. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevSTPER.8.010107

[6] W. Blasiak, M. Godlewska, R. Rosiek, and D. Wcislo, "Spectrum of physics comprehension," *European Journal of Physics*, vol. 33, p. 565, 2012. [Online]. Available: https://www.researchgate.net/publication/254496494_Spectrum_of_physics_comprehension

[7] W. Tenachi, R. Ibata, and F. I. Diakogiannis, "Deep symbolic regression for physics guided by units constraints: Toward the automated discovery of physical laws," p. 99, dec 2023. [Online]. Available: https://dx.doi.org/10.3847/1538-4357/ad014c

[8] Z.-Y. Khoo, A. Yang, J. S. C. Low, and S. Bressan, "Celestial machine learning: From data to mars and beyond with ai feynman," in *Database and Expert Systems Applications: 34th International Conference, DEXA 2023, Penang, Malaysia, August 28–30, 2023, Proceedings, Part II*. Berlin, Heidelberg: Springer-Verlag, 2023, p. 469–474. [Online]. Available: https://doi.org/10.1007/978-3-031-39821-6_41

[9] J. Kepler, *Astronomia Nova*, 1609, reviewed in Nature Journal, URL: https://www.nature.com/articles/462725a. [Online]. Available: https://muse.jhu.edu/book/81252

[10] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics: The New Millennium Edition: Mainly Mechanics, Radiation, and Heat.* Basic Books, 1963, vol. 1. [Online]. Available: https://books.google.com/books?id=d76DBQAAQBAJ

[11] ——, *The Feynman Lectures on Physics*, ser. The Feynman Lectures on Physics. Pearson/Addison-Wesley, 1963, vol. 3. [Online]. Available: https://books.google.com/books?id=_6XvAAAAMAAJ

[12] ——, *The Feynman Lectures on Physics*, ser. The Feynman Lectures on Physics. Pearson/Addison-Wesley, 1963, vol. 2. [Online]. Available: https://books.google.com/books?id=AbruAAAAMAAJ

[13] B. K. Petersen, M. L. Larma, T. N. Mundhenk, C. P. Santiago, S. K. Kim, and J. T. Kim, "Deep symbolic regression: Recovering mathematical expressions from data via risk-seeking policy gradients," in *International Conference on Learning Representations*, 2021. [Online]. Available: https://openreview.net/forum?id=m5Qsh0kBQG

[14] N. Makke and S. Chawla, "Interpretable scientific discovery with symbolic regression: a review," vol. 57, no. 1, 2024, p. 2. [Online]. Available: https://doi.org/10.1007/s10462-023-10622-0

[15] K. S. Taber, "Maths should be the last thing we teach," *Physics Education*, vol. 44, no. 4, p. 336, jul 2009. [Online]. Available: https://dx.doi.org/10.1088/0031-9120/44/4/F01

[16] L. J. W. S. Rayleigh, *The Theory of Sound.* Dover Publications, 1915, reprinted edition. [Online]. Available: https://shorturl.at/kAUZ2

[17] Y. Liang, D. Zou, H. Xie, and F. L. Wang, "Exploring the potential of using chatgpt in physics education," *Smart Learning Environments*, vol. 10, no. 1, p. 52, 2023. [Online]. Available: https://doi.org/10.1186/s40561-023-00273-7

[18] F. Kieser, P. Wulff, J. Kuhn, and S. Küchemann, "Educational data augmentation in physics education research using chatgpt," *Phys. Rev. Phys. Educ. Res.*, vol. 19, p. 020150, Oct 2023. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevPhysEducRes.19.020150

[19] A. Birhane, A. Kasirzadeh, D. Leslie, and S. Wachter, "Science in the age of large language models," *Nature Reviews Physics*, vol. 5, pp. 277–280, 2023. [Online]. Available: https://www.nature.com/articles/s42254-023-00581-4

[20] D. Gerwin, "Information processing, data inferences, and scientific generalization," *Behavioral Science*, vol. 19, no. 5, pp. 314–325, 1974. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/bs.3830190504

[21] M. Virgolin and S. P. Pissis, "Symbolic regression is NP-hard," *Transactions on Machine Learning Research*, 2022. [Online]. Available: https://openreview.net/forum?id=LTiaPxqe2e

[22] M. Willis, H. Hiden, and G. Montague, "Developing inferential estimation algorithms using genetic programming," *IFAC Proceedings Volumes*, vol. 30, no. 9, pp. 209–214, 1997, iFAC Symposium on Advanced Control of Chemical Processes 1997 (ADCHEM '97), Banff, Canada, 9-11 June. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1474667017431624

[23] J. H. Holland, *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence.* The MIT Press, 1975, in Special Collection: CogNet. [Online]. Available: https://mitpress.mit.edu/9780262581110/adaptation-in-natural-and-artificial-systems/

[24] N. L. Cramer, "A representation for the adaptive generation of simple sequential programs," in *Proceedings of the First International Conference on Genetic Algorithms*, 1985, pp. 183–187. [Online]. Available: https://dl.acm.org/doi/10.5555/645511.657085

[25] J. R. Koza, "Genetic programming as a means for programming computers by natural selection," *Statistical Computing*, vol. 4, pp. 87–112, 1994. [Online]. Available: https://link.springer.com/article/10.1007/BF00175355

[26] L. Biggio, T. Bendinelli, A. Neitz, A. Lucchi, and G. Parascandolo, "Neural symbolic regression that scales," in *Proceedings of the 38th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. Meila and T. Zhang, Eds., vol. 139. PMLR, 18–24 Jul 2021, pp. 936–945. [Online]. Available: https://proceedings.mlr.press/v139/biggio21a.html

[27] J. Kubalík, E. Derner, and R. Babuška, "Toward physically plausible data-driven models: A novel neural network approach to symbolic regression," *IEEE Access*, vol. 11, pp. 61481–61501, 2023. [Online]. Available: https://arxiv.org/abs/2302.00773

[28] G. Chen, "Learning symbolic expressions via gumbel-max equation learner network," *arXiv preprint arXiv:2012.06921*, 2020. [Online]. Available: https://arxiv.org/abs/2012.06921

[29] M. Zhang, S. Kim, P. Y. Lu, and M. Soljačić, "Deep learning and symbolic regression for discovering parametric equations," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–13, 2023.

[30] S.-M. Udrescu and M. Tegmark, "Ai feynman: a physics-inspired method for symbolic regression," *Science Advances*, vol. 6, no. 16, p. eaay2631, 2020. [Online]. Available: https://www.science.org/doi/abs/10.1126/sciadv.aay2631

[31] L. S. Keren, A. Liberzon, and T. Lazebnik, "A computational framework for physics-informed symbolic regression with straightforward integration of domain knowledge," *Scientific Reports*, vol. 13, no. 1, p. 1249, 2023. [Online]. Available: https://doi.org/10.1038/s41598-023-28328-2

[32] K. M. Galotti, "Approaches to studying formal and everyday reasoning," *Psychological Bulletin*, vol. 105, no. 3, pp. 331–351, 1989. [Online]. Available: https://psycnet.apa.org/record/1989-28293-001

[33] J. White, Q. Fu, S. Hays, M. Sandborn, C. Olea, H. Gilbert, A. Elnashar, J. Spencer-Smith, and D. C. Schmidt, "A prompt pattern catalog to enhance prompt engineering with chatgpt," *ArXiv*, vol. abs/2302.11382, 2023. [Online]. Available: https://api.semanticscholar.org/CorpusID:257079092

[34] D. Rajagopal, V. Khetan, B. Sacaleanu, A. Gershman, A. E. F. Fano, and E. Hovy, "Template filling for controllable commonsense reasoning," in *Findings of the Association for Computational Linguistics: IJCNLP-AACL 2023 (Findings)*. Nusa Dua, Bali: Association for Computational Linguistics, Nov. 2023, pp. 250–260. [Online]. Available: https://aclanthology.org/2023.findings-ijcnlp.23

[35] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. H. Chi, Q. V. Le, and D. Zhou, "Chain-of-thought prompting elicits reasoning in large language models," in *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022. [Online]. Available: http://papers.nips.cc/paper_files/paper/2022/hash/9d5609613524ecf4f15af0f7b31abca4-Abstract-Conference.html

[36] Y. Lu, M. Bartolo, A. Moore, S. Riedel, and P. Stenetorp, "Fantastically ordered prompts and where to find them: Overcoming few-shot prompt order sensitivity," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Dublin, Ireland: Association for Computational Linguistics, May 2022, pp. 8086–8098. [Online]. Available: https://aclanthology.org/2022.acl-long.556

[37] A. Webson and E. Pavlick, "Do prompt-based models really understand the meaning of their prompts?" in *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Seattle, United States: Association for Computational Linguistics, Jul. 2022, pp. 2300–2344. [Online]. Available: https://aclanthology.org/2022.naacl-main.167

[38] M. Kazemi, N. Kim, D. Bhatia, X. Xu, and D. Ramachandran, "LAMBADA: Backward chaining for automated reasoning in natural language," in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Toronto, Canada: Association for Computational Linguistics, Jul. 2023, pp. 6547–6568. [Online]. Available: https://aclanthology.org/2023.acl-long.361

[39] A. Creswell and M. Shanahan, "Faithful reasoning using large language models," *arXiv preprint arXiv:2208.14271*, 2022. [Online]. Available: https://arxiv.org/abs/2208.14271

[40] O. Press, M. Zhang, S. Min, L. Schmidt, N. A. Smith, and M. Lewis, "Measuring and narrowing the compositionality gap in language models," 2023. [Online]. Available: https://openreview.net/forum?id=PUwbwZJz9dO

[41] J. Liu, A. Liu, X. Lu, S. Welleck, P. West, R. L. Bras, Y. Choi, and H. Hajishirzi, "Generated knowledge prompting for commonsense reasoning," 2022.

[42] D. Zhou, N. Schärli, L. Hou, J. Wei, N. Scales, X. Wang, D. Schuurmans, C. Cui, O. Bousquet, Q. V. Le, and E. H. Chi, "Least-to-most prompting enables complex reasoning in large language models," in *The Eleventh International Conference on Learning Representations*, 2023. [Online]. Available: https://openreview.net/forum?id=WZH7099tgfM

[43] A. Madaan, N. Tandon, P. Gupta, S. Hallinan, L. Gao, S. Wiegreffe, U. Alon, N. Dziri, S. Prabhumoye, Y. Yang, S. Gupta, B. P. Majumder, K. Hermann, S. Welleck, A. Yazdanbakhsh, and P. Clark, "Self-refine: Iterative refinement with self-feedback," in *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. [Online]. Available: https://openreview.net/forum?id=S37hOerQLB