A Multi-agent Framework for Physical Laws Discovery

Bo Hu [1] **Siyu Liu** [1 2] **Beilin Ye** [1] **David J. Srolovitz** [1 2] **Tongqi Wen** [1 2]
[1] *Center for Structural Materials, Department of Mechanical Engineering, The University of Hong Kong, Hong Kong, China* [2] *Materials Innovation Institute for Life Sciences and Energy (MILES), HKU-SIRI, Shenzhen, China*. Correspondence to: Tongqi Wen tongqwen@hku.hk.

1. Introduction

Discovering explicit physical laws has traditionally depended on human intuition and domain expertise. Recent advances in artificial intelligence, particularly large language models (LLMs), offer a new route to accelerate this process by automating key steps from hypothesis generation to interpretable model construction. Here we develop an LLM-based multi-agent framework for physical-law discovery that integrates literature-guided variable selection, hypothesis formulation, symbolic regression, formula derivation, and mechanistic explanation. We validate the framework on three representative materials problems: the glass-forming ability (GFA) of metallic glasses, the Vickers hardness of compounds, and the Young's modulus of multi-component alloys. Using physically and chemically meaningful descriptors as inputs, the discovered formulas achieve strong agreement with reference data, with correlation coefficients up to 0.94 (GFA), 0.86 (hardness), and 0.94 (Young's modulus), while remaining compact and interpretable. Beyond fitting, the Young's modulus formula generalizes to quaternary and quinary alloys, improving prediction accuracy by up to 78.8% relative to the classical rule of mixtures. By integrating cross-disciplinary knowledge, reflection mechanisms, and expert-like reasoning ability into symbolic regression, our AI-centric framework offers a robust and extensible platform for automated physical laws discovery, demonstrating that AI can increasingly serve as an essential role in modern scientific research by thinking and acting like field experts.

2. Substantial section

We introduce a general LLM-driven multi-agent framework [1, 2] for the automated discovery of physical laws from scientific data. The framework orchestrates multiple agents to emulate core stages of the scientific workflow, including literature review, hypothesis generation, data curation, symbolic regression, and the derivation and interpretation of closed-form equations (Fig. 1). Methodologically, our symbolic regression module combines beam search with LLM-based reflection mechanisms [3, 4] to propose, evaluate, and iteratively refine candidate expressions in a structured manner. We evaluate the framework on three representative materials-science problems: predicting the glass-forming ability (GFA) of metallic glasses, the Vickers hardness of compounds [5], and the Young's modulus of multi-principal element alloys (MPEAs) [6]. Using Gemini-2.5-flash as the base model, the discovered formulas achieve predictive performance of up to $R^2 = 0.94$, 0.86, and 0.94 on these tasks, respectively (Table 1). Because experimental materials data can be noisy and heterogeneous, predictive metrics alone may not fully reflect the scientific utility of the inferred laws. To further assess practical value and generalization, we apply the discovered Young's modulus equation to the design of previously unseen quaternary and quinary MPEAs. The resulting formula reduces the mean absolute percentage error (MAPE) by up to 78.8% relative to existing empirical relations, while offering substantially higher computational efficiency than first-principles calculations and foundation atomic models. Together, these results demonstrate that LLM-enabled multi-agent reasoning can bridge data-driven learning and interpretable symbolic modeling, providing a scalable route toward automated, physically meaningful law discovery across scientific domains.

2.1 Related work

Scientific progress is often marked by the discovery of compact physical laws–from Newton's laws of motion to Kepler's laws of planetary dynamics [7]–typically expressed as explicit mathematical equations that expose underlying mechanisms. Such laws not only summarize observations but also enable explanation, extrapolation, and the formation of new conceptual frameworks; indeed, paradigm shifts frequently co-evolve with the articulation of new governing principles [8]. Although modern deep learning models can deliver highly accurate predictions, their limited interpretability hampers the extraction of mechanistic insight and generalizable scientific understanding. In contrast, closed-form expressions directly encode relationships among variables and therefore provide a natural basis for interpretation and theory building. Consequently, inferring interpretable equations from experimental and simulation data remains a central challenge in data-driven science. Symbolic regression (SR) has emerged as a leading explainable approach for this purpose, aiming to identify analytical expressions that best describe the dependencies among physical variables [9, 10]. More recently, LLMs, leveraging extensive pre-training and strong priors over scientific syntax and concepts, have been incorporated to guide and accelerate SR, thereby improving the automated discovery of governing equations [11, 12, 13]. However, an end-to-end pipeline for discovering physical laws, spanning literature-based data collection, hypothesis generation, symbolic formula derivation, evaluation, and interpretation, remains limited.

2.2 *Figures and tables*

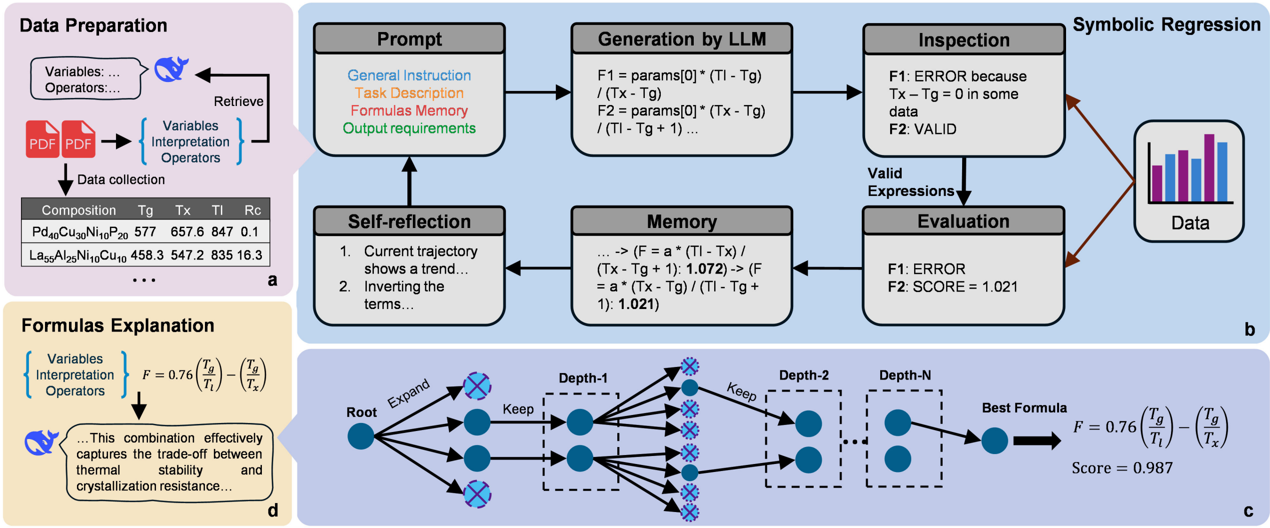

Fig. 1: Schematic of the proposed multi-agent framework for physical-law discovery. (a) A reasoning LLM conducts literature review and supports data preparation. (b) Multi-agent collaboration for proposing, evaluating, and refining candidate formulas with trajectory-based feedback. (c) Beam search balances predictive accuracy and expression complexity by retaining the top-K candidates at each depth. (d) An explanation agent with RAG contextualizes and interprets the discovered formulas.

Table 1: Average predictive performance of *DeepSeek-V3* and *Gemini-2.5-flash* across the considered tasks. Metrics are reported as "train / test" for the root-mean-square error (RMSE) and coefficient of determination (R^2). Best values are highlighted in *italic*.

Task	DeepSeek-V3		Gemini-2.5-flash	
	RMSE	R^2	RMSE	R^2
GFA (log[K/s])	2.43 / 2.41	0.80 / 0.82	*1.92 / 1.52*	*0.89 / 0.94*
Hardness (GPa)	5.05 / 4.86	0.84 / 0.84	*4.90 / 4.52*	*0.85 / 0.86*
Young's modulus (GPa)	40.73 / 39.86	0.72 / 0.73[a]	*21.50 / 20.75*	*0.94 / 0.94*

[a] For the Young's modulus task, one of the three *DeepSeek-V3* runs failed to produce a competitive formula, leading to a markedly lower average. The three runs yielded R^2 values of 0.94/0.95, 0.36/0.36, and 0.87/0.89 (train/test), respectively.

Acknowledgments

The work described is partially supported by a grant from the NSFC/RGC Joint Research Scheme sponsored by the Research Grants Council of the Hong Kong Special Administrative Region, China and the National Natural Science Foundation of China (Project No. N_HKU767/25). The authors would like to thank for startup funding from Materials Innovation Institute for Life Sciences and Energy (MILES), HKU-SIRI in Shenzhen for support of this manuscript. This work is partially supported by Research Grants Council, Hong Kong SAR through the General Research Fund (17210723, 17200424). T. W. acknowledges additional support by The University of Hong Kong (HKU) via seed funds (2509100468) and Guangdong Natural Science Fund (2025A1515012129).

References

[1] Qingyun Wu, Gagan Bansal, Jieyu Zhang, Yiran Wu, Beibin Li, Erkang Zhu, Li Jiang, Xiaoyun Zhang, Shaokun Zhang, Jiale Liu, Ahmed Hassan Awadallah, Ryen W White, Doug Burger, and Chi Wang. Autogen: Enabling next-gen LLM applications via multi-agent conversations. In *First Conference on Language Modeling*, 2024.

[2] Guohao Li, Hasan Hammoud, Hani Itani, Dmitrii Khizbullin, and Bernard Ghanem. Camel: Communicative agents for "mind" exploration of large language model society. In *Advances in Neural Information Processing Systems*, volume 36, pages 51991–52008, 2023.

[3] Noah Shinn, Federico Cassano, Ashwin Gopinath, Karthik Narasimhan, and Shunyu Yao. Reflexion: language agents with verbal reinforcement learning. In *Advances in Neural Information Processing Systems*, volume 36, pages 8634–8652, 2023.

[4] Zhibin Gou, Zhihong Shao, Yeyun Gong, Yelong Shen, Yujiu Yang, Nan Duan, and Weizhu Chen. CRITIC: Large Language Models Can Self-Correct with Tool-Interactive Critiquing. *arXiv:2305.11738*, 2024.

[5] Christian Tantardini, Hayk A. Zakaryan, Zhong-Kang Han, Tariq Altalhi, Sergey V. Levchenko, Alexander G. Kvashnin, and Boris I. Yakobson. Material Hardness Descriptor Derived by Symbolic Regression. *Journal of Computational Science*, 82:102402, 2024.

[6] Guillermo Vazquez, Prashant Singh, Daniel Sauceda, Richard Couperthwaite, Nicholas Britt, Khaled Youssef, Duane D. Johnson, and Raymundo Arróyave. Efficient machine-learning model for fast assessment of elastic properties of high-entropy alloys. *Acta Materialia*, 232:117924, 2022.

[7] Silviu-Marian Udrescu and Max Tegmark. AI Feynman: A physics-inspired method for symbolic regression. *Science Advances*, 6(16):eaay2631, 2020.

[8] Thomas S Kuhn. *The structure of scientific revolutions*, volume 962. University of Chicago press Chicago, 1997.

[9] Hanchen Wang, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, Shengchao Liu, Peter Van Katwyk, Andreea Deac, Anima Anandkumar, Karianne Bergen, Carla P. Gomes, Shirley Ho, Pushmeet Kohli, Joan Lasenby, Jure Leskovec, Tie-Yan Liu, Arjun Manrai, Debora Marks, Bharath Ramsundar, Le Song, Jimeng Sun, Jian Tang, Petar Veličković, Max Welling, Linfeng Zhang, Connor W. Coley, Yoshua Bengio, and Marinka Zitnik. Scientific discovery in the age of artificial intelligence. *Nature*, 620(7972):47–60, 2023.

[10] Dimitrios Angelis, Filippos Sofos, and Theodoros E. Karakasidis. Artificial Intelligence in Physical Sciences: Symbolic Regression Trends and Perspectives. *Archives of Computational Methods in Engineering*, 30(6):3845–3865, 2023.

[11] Matteo Merler, Katsiaryna Haitsiukevich, Nicola Dainese, and Pekka Marttinen. In-context symbolic regression: Leveraging large language models for function discovery. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop)*, pages 427–444, 2024.

[12] Parshin Shojaee, Kazem Meidani, Shashank Gupta, Amir Barati Farimani, and Chandan K. Reddy. LLM-SR: Scientific Equation Discovery via Programming with Large Language Models. *arXiv:2404.18400*, 2024.

[13] Mengge Du, Yuntian Chen, Zhongzheng Wang, Longfeng Nie, and Dongxiao Zhang. LLM4ED: Large Language Models for Automatic Equation Discovery. *arXiv:2405.07761*, 2024.