Geometry-Informed AI for Scientific Discovery

Melanie Weber

Artificial intelligence offers tremendous opportunities to accelerate scientific discovery. Already, AI models have driven major breakthroughs in areas such as protein folding and weather forecasting. Going forward, AI may hold the keys to the development of treatments for presently incurable diseases, to the design of innovative new materials, and to resolving decades-long open problems in mathematics. Currently, the model landscape is dominated by transformers; however, their extreme data and compute requirements and limited interpretability present fundamental challenges for the road ahead. Progress toward the next generation of AI for science will require a shift toward smaller, *structured* models that are efficient, interpretable, and capable of acting as genuine scientific collaborators: guiding experiments, revealing hidden patterns in data, and helping to chart new scientific frontiers.

One possible avenue for developing such next-generation models, which I am pursuing in my research, is *geometry-informed* AI.[1] This is motivated by the observation that encoding data geometry as inductive bias into models can mitigate high resource demands by reducing the model's data and computing needs, as demonstrated by recent work from my group and others.[2] To illustrate this, consider the example of image classification, in which labels are typically assigned based on the objects shown in the image, irrespective of their location. One can encode such structure into a model by only considering model architectures whose outputs are agnostic to shifts of objects within the image. Convolutional neural networks, which achieved a major breakthrough in image classification in the 2010s, are an early example of a model architecture that has this property.[3] More broadly, models can encode a variety of geometric structures, such as symmetries arising from fundamental laws of physics or low-dimensional structure reflecting inherent correlations in high-dimensional data.

Why does this matter for the sciences? Take state-of-the-art weather prediction models, which require training on millions of examples to provide accurate forecasts. By encoding known geometric structure, we can prevent models from expending resources on implausible scenarios. This is accomplished by constrain-

https://doi.org/10.1162/DAED.a.992

ing model outputs to remain consistent with established domain knowledge, such as physical laws that enforce specific symmetries.

Geometric models are already driving progress across scientific domains: Equivariant models accelerate materials discovery and physics-informed architectures achieve high-resolution weather forecasting.[4] In my own interdisciplinary work, we have developed geometric approaches for representing large-scale single-cell data in a way that makes it easier for researchers to trace how cells mature and branch into different types in developmental processes.[5] For astronomy, we developed a method that computes interpretable representations of galaxies that preserve information about their surrounding cosmic environment, uncovering patterns tied to galaxy mass and star formation.[6] In structural biology, we introduced a geometry-aware framework for heterogeneous cryo-EM reconstruction that predicts atomic backbone conformations, allowing researchers to recover protein structural variability from single-particle images using geometric priors.[7]

State-of-the-art geometric models can already rival transformer-based architectures in specialized tasks, yet they represent only an early glimpse of what geometry-informed AI could enable. To move beyond specialized solutions and toward more versatile AI assistants, both mathematics and engineering advances are needed. Recent theoretical work by my group and others has begun to clarify when and why geometric priors yield provable advantages, but deeper mathematical foundations, and their incorporation into broadly applicable foundation models, will be key to accelerating progress in geometry-informed AI.[8] Equally important is the development of architectures that encode richer and more heterogeneous geometric structure, along with the software and hardware infrastructure needed to scale them efficiently.[9] Together, these efforts could enable geometric models capable of driving major advances in AI-assisted scientific discovery.

While I have emphasized that mathematical insights can advance artificial intelligence, the converse is increasingly true as well. AI-assisted reasoning tools are being introduced into mathematical research, with early successes showing that they are particularly effective for "needle in a haystack" problems, where a correct construction, counterexample, or identity exists but is extremely difficult to find with classical approaches.[10] Since mathematics is fundamentally concerned with identifying and exploiting structure, I expect that structured models, such as geometry-informed AI, will play an important role going forward. Such approaches could mitigate some of the growing pains of current methods: They rely heavily on large language models that require time-consuming human formalization of mathematical data as text or draw on unreliable informal sources. Their outputs are also difficult for humans to interpret, limiting the extraction of proof strategies and mathematical intuition. Encoding mathematical principles directly into models could enable them to reason in ways that more closely align with human intuition, strengthening human-AI collaboration in research mathematics. In my own research, we have already seen early successes with geometric models in combinatorics.[11]

Looking ahead, I envision a scientific ecosystem in which AI is not merely a tool but a collaborator. While current models primarily assist with data analysis, future models could help researchers interrogate existing literature, act as sounding boards for hypothesis generation, and collaborate in the conception of new methodologies. Such human-AI collaborations have the potential to accelerate progress toward answering the great scientific questions of our time, and those yet to emerge.

ABOUT THE AUTHOR

Melanie Weber is an Assistant Professor of Applied Mathematics and of Computer Science at Harvard University, where she leads the Geometric Machine Learning Group. She is also a 2025 Early Career Fellow with the Schmidt Sciences AI2050 Project and a 2024 Research Fellow in Mathematics with the Alfred P. Sloan Foundation. Her research studies geometric structure in data and models and how to leverage such information for the design of new, efficient machine-learning methods with provable guarantees.

ENDNOTES

1 Melanie Weber, "Geometric Machine Learning," *AI Magazine*, January 10, 2025, https://doi.org/10.1002/aaai.12210.

2 Alberto Bietti, Luca Venturi, and Joan Bruna, "On the Sample Complexity of Learning Under Invariance and Geometric Stability," in *NeurIPS '21: Proceedings of the 35th International Conference on Neural Information Processing Systems*, ed. Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, et al. (Curran Associates, Inc., 2021); Song Mei, Theodor Misiakiewicz, and Andrea Montanari, "Learning with Invariances in Random Features and Kernel Models," *Proceedings of Machine Learning Research* 134 (2021); Bobak Kiani, Thien Le, Hannah Lawrence, et al., "On the Hardness of Learning Under Symmetries," paper presented at the Twelfth International Conference on Learning Representations, Vienna, Austria, May 7, 2024, https://openreview.net/forum?id=ARPrtuzAnQ; Bobak T. Kiani, Jason Wang, and Melanie Weber, "On the Hardness of Learning Neural Networks Under the Manifold Hypothesis," in *NeurIPS '24: Proceedings of the 38th International Conference on Neural Information Processing Systems*, ed. Amir Globerson, Lester Mackey, Danielle Belgrave, et al. (Curran Associates, Inc., 2024); and Johann Brehmer, Sönke Behrends, Pim De Haan, and Taco Cohen, "Does Equivariance Matter at Scale?" *Transactions on Machine Learning Research* (2025), https://openreview.net/pdf?id=wilNute8Tn.

3 Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner, "Gradient-Based Learning Applied to Document Recognition," *Proceedings of the IEEE* 86 (11) (1998).

4 Simon Batzner, Albert Musaelian, Lixin Sun, et al., "E(3)-Equivariant Graph Neural Networks for Data-Efficient and Accurate Interatomic Potentials," *Nature Communications* 13

(2022); and Jaideep Pathak, Shashank Subramanian, Peter Harrington, et al., "FourCastNet: A Global Data-Driven High-Resolution Weather Model Using Adaptive Fourier Neural Operators," arXiv (2022), https://doi.org/10.48550/arXiv.2202.11214.

[5] Nithya Bhasker, Hattie Chung, Louis Boucherie, et al., "Uncovering Developmental Lineages from Single-Cell Data with Contrastive Poincaré Maps," bioRxiv (2025), https://doi.org/10.1101/2025.08.22.671789.

[6] Ana Sofia Uzoy, Claire Lamman, and Melanie Weber, "Manifold Learning for Cosmic Structures," paper presented at the Thirty-Ninth Annual Conference on Neural Information Processing Systems (NeurIPS 2025), San Diego, California, December 1, 2025, https://neurips.cc/virtual/2025/loc/san-diego/122903.

[7] Jonathan Krook, Axel Janson, Joakim Andén, et al., "Protein Graph Neural Networks for Heterogeneous Cryo-EM Reconstruction," paper presented at the IEEE International Conference on Image Processing, Tampere, Finland, September 13–17, 2026, https://2026.ieeeicip.org.

[8] Bietti, Venturi, and Bruna, "On the Sample Complexity of Learning Under Invariance and Geometric Stability"; Mei, Misiakiewicz, and Montanari, "Learning with Invariances in Random Features and Kernel Models"; Kiani, Le, Lawrence, et al., "On the Hardness of Learning Under Symmetries"; and Kiani, Wang, and Weber, "On the Hardness of Learning Neural Networks Under the Manifold Hypothesis."

[9] Neil He, Jiahong Liu, Buze Zhang, et al., "Position: Beyond Euclidean–Foundation Models Should Embrace Non-Euclidean Geometries," paper presented at Learning on Graphs Conference 2025, Arizona State University, December 10, 2025, https://openreview.net/forum?id=WoK409olln.

[10] François Charton, Jordan S. Ellenberg, Adam Zsolt Wagner, and Geordie Williamson, "Patternboost: Constructions in Mathematics with a Little Help from AI," arXiv (2024), https://doi.org/10.48550/arXiv.2411.00566; Alex Davies, Petar Veličković, Lars Buesing, et al., "Advancing Mathematics by Guiding Human Intuition with AI," *Nature* 600 (7887) (2021); and Alexander Novikov, Ngân Vũ, Marvin Eisenberger, et al., "AlphaEvolve: A Coding Agent for Scientific and Algorithmic Discovery," arXiv (2025), https://doi.org/10.48550/arXiv.2506.13131.

[11] Knut Vanderbush and Melanie Weber, "Neural Algorithmic Reasoning for Approximate *k*-Coloring with Recursive Warm Starts," arXiv (2026), https://arxiv.org/html/2601.05137v1.