MIT Open Access Articles

Artificial intelligence in computational materials science

Massachusetts Institute of Technology

Artificial intelligence in computational materials science: Special issue overview

Artificial Intelligence in computational materials science: special issue overview

Heather J. Kulik[1] and Pratyush Tiwary[2]*

[1]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, United States

[2]Department of Chemistry and Biochemistry and Institute for Physical Science and Technology, University of Maryland, College Park 20742, United States; Email: ptiwary@umd.edu

Keywords: artificial intelligence, materials science, data sciences

* Contributing author email: ptiwary@umd.edu

Abstract: In this themed collection we aim to broadly review some of the critical, recent progress in the application of AI/ML to various aspects of computational materials science and materials science more broadly. In this themed collection spread across two issues, we have assembled a collection of articles from leaders in the broad domain of applying AI/ML, which we collectively refer to as ML, in computational materials science. Together these articles curate the critical, recent progress in the application of ML to various aspects of materials science. These include ML approaches for understanding and driving electron microscopy, designing energy materials and the discovery of principles and materials relevant to the design of materials for the future, studying crystal nucleation and growth, the use of ML to describe force-fields governing material and molecular behavior and other topics.

The materials science community across academia, national laboratories and industry has long benefitted from and contributed to the development of increasingly sophisticated quantitative methods. More recently, data-driven approaches for materials science have seen a heavy use of Machine Learning (ML) and Artificial Intelligence (AI) based ideas. These approaches have made it possible to reveal predictive patterns in the triad of structure–property-function relationships across all branches of materials sciences and engineering. The synergistic interactions between materials science and data sciences continue to flourish with tremendous advances in computing power, software, and algorithms, as well as enormous increases in data available from experiment and simulation. Perhaps it is hard to say which of the two has expanded more in the last decade - the amount of data available through increasingly sophisticated simulations and experiments, or, the sophistication of the ML/AI algorithms available to make sense of the data. One has also witnessed robust developments where data sciences and materials experimentation are no longer two separate entities but instead deeply integrated. In these so-called active learning paradigms, data driven approaches are used to guide further experiments, often in a closed loop iterative manner. Such autonomous materials discovery

paradigms are enabling navigating material space in a high throughput yet efficient and controlled manner.

The impact of AI/ML is not limited to just smarter combinatorial design of materials. It is now also possible to train AI and ML models to gain fundamental understanding and extract patterns at spatiotemporal scales that were previously impossible with conventional computational materials modeling or with the best available theories. Powerful open-source toolkits for AI/ML model training and architecture selection have also made ML more accessible to researchers with diverse training backgrounds. As a result, in laboratories across the world, scientists and engineers are identifying ways in which they can incorporate AI and ML into their research.

Nevertheless, some key challenges remain in the application of AI and ML to materials discovery. While the text-based extraction of prior experimental efforts for data-driven models is maturing [1], progress is only just beginning in the systematic analysis of experimental images [2-4]. Furthermore, positive publishing bias means that AI models are challenged to find good sources of failed experiments. Work has indicated the benefit of these failed experiences to inform AI models and that human bias can influence extracted predictions[5-6]. Creative strategies have been devised to simulate negative data for classification tasks [7]. Within the context of ML for the acceleration of physics based modeling, a key outstanding challenge is the quality of the data source. For example, many materials prediction models are trained on approximate density functional theory (DFT), inheriting the bias of the underlying functional. While ML-derived functionals represent one emerging strategy to overcome limitations in DFT [8], there remains no one-size-fits-all physics based model that is established to be predictive across materials space. Thus, strategies that incorporate the uncertainty of the physics-based method [9] or make recommendations [10] about the most suitable method to employ are needed. There is indeed early evidence that AI tools can also support experts in their choice of physics-based methodology. In this issue, we summarize key ways researchers are advancing AI in spite of these potential challenges.

AI methods are increasingly useful in interpreting, analyzing and complementing the static and dynamic datasets generated from different spectroscopic and microscopic techniques. In this connection, Kalinin and colleagues [11] as well and Chan and colleagues [12] review the state-of-the-art in how ML can be used to better understand the huge amounts of data being generated in electron and scanning probe microscopy techniques. Their articles show how such understandings can be complemented with theory, and leveraged in a closed loop manner to perform automated microscopy experiments and eventually open the path towards direct atomic fabrication with active learning augmented microscopy.

The contribution by Takeuchi, Kusne and co-workers [13] looks at yet another aspect of how ML and specifically active learning is revolutionizing autonomous materials design. They show how ML can be used for autonomous model exploration relevant to materials discovery. They also consider the very relevant question of ML related education for physical sciences students and report efforts they have undertaken to educate a large

number of physical sciences and engineering students in ML. Their efforts have involved the very creative development and use of the so-called LEGOLAS education kit – a LEGO®* based Low-cost Autonomous Scientist.

Articles from the groups of Viswanathan [14] and Mannodi-Kanakkithodi [15] consider the broad question of high throughput materials design using ML. The contribution from Viswanathan and co-workers describes an automated workflow named AutoMat for the automated discovery of electrochemical materials critical to large-scale electrification. AutoMat accelerates the computational steps fundamental to such materials design through a variety of advances in theory, software engineering and machine learning. The contribution from Mannodi-Kanakkithodi and group overviews high-throughput computations and ML methods for the very important problem of Halide Perovskite discovery. They examine specific approaches that make it possible to predict in an accelerated manner various material properties and screen through enormous chemical spaces. The key approaches they describe involve a tight integration of Density Functional Theory (DFT) simulations with ML.

The article by Ceriotti [16] examines very carefully the future of simulations itself. It discusses how ML based theoretical methods that will enable molecular simulations with quantum level accuracy, yet with the cost of performing classical simulations. The article highlights various developments that have made ML based interatomic potentials a viable option for materials simulations. At the same time the article also makes a case for how more sophisticated electronic structure calculations will be continually needed as we push the boundaries of computational materials science and what we aim to achieve through it.

ML methods are starting to automate and revolutionize the fields of crystal structure prediction as well as predicting stable polymorphs, their free energies and kinetics. In this context, we have submissions from the Day [17] group as well as the groups of Rogal and Sarupria [18]. Day and collaborators do a deep dive into the problem of crystal structure prediction with machine learning methods. They cover progress made and challenges remaining in multiple aspects of crystal structure prediction, including the evaluation of accurate energies, mapping the structural landscape and inverse design of molecules given a target property in mind.

Fundamental to crystallization is the process of nucleation, which is a prototypical rare event that cannot be simulated using classical molecular dynamics as is even with the fastest available supercomputers. Rogal and Sarupria present the state-of-the-art in performing specialized yet robust molecular simulations that can reach the experimentally relevant timescales for nucleation and growth processes. These types of simulations are now making it possible to directly observe nucleation in all-atom resolution of generic systems, and also gain insight into the reaction coordinate or driving forces behind the nucleation.

Arguably deeply connected with the process of crystallization is that of self-assembly, albeit often with different length scales and driving forces. This term is loosely used for the process in which different systems' constituents organize themselves into highly

ordered structures. Huang and colleagues describe [19] the use of machine learning algorithms for the development of kinetic network models (KNMs) that aid interpretation of MD simulations. Huang and colleagues describe how KNMs can capture self-assembly processes such as crystallization of colloidal particles. They also describe an outlook where increasingly deep learning algorithms such as graph neural networks can be exploited to understand and interpret these self-assembly events from complex trajectories.

To conclude, the articles these two special issues present an overview of some of the challenges the computational and broader materials science community is being able to solve with recent AI/ML methods. The articles also discuss numerous open problems and avenues for future research, hinting at a scientific discipline which has a vibrant, active future ahead. Naturally, we cannot span the entirety of the fields in a limited number of articles and thus in no way are these special issues meant to be complete. Just to name a few, an important topic we did not cover here is the fair, open and equitable access to data for training ML models – we refer to this recent excellent overview instead [20]. A second important topic not considered here was the interpretation of AI based models in materials science [21]. Widely used theories such as classical nucleation theory or the Allen-Cahn equation might have their approximations. However, they provide one not just predictive power but also intuition into the underlying physics and chemistry. AI and ML tools arguably provide more predictive power than these revered theories, but often this comes at the cost of understanding. By developing interpretable AI models perhaps this gap can be bridged. We hope the collection of articles will be found useful by the materials science community.

Conflict of interest and data availability statement

On behalf of all authors, the corresponding author states that there is no conflict of interest. No data was used in writing this review article.

References

[1] Kononova O, He T, Huo H, Trewartha A, Olivetti EA, Ceder G. Opportunities and Challenges of Text Mining in Materials Research. *iScience* 2021, **24**:102155 http://dx.doi.org/10.1016/j.isci.2021.102155

[2] Jiang W, Schwenker E, Spreadbury T, Ferrier N, Chan MKY, Cossairt O. A Two-stage Framework for Compound Figure Separation. *arXiv* 2021, http://dx.doi.org/arXiv:2101.09903

[3] Schwenker E, Jiang W, Spreadbury T, Ferrier N, Cossairt O, Chan MKY: EXSCLAIM! -- An Automated Pipeline for the Construction of Labeled Materials Imaging Datasets from Literature. *arXiv* 2021, http://dx.doi.org/arXiv:2103.10631

[4] Mukaddem KT, Beard EJ, Yildirim B, Cole JM: ImageDataExtractor: A Tool To Extract and Quantify Data from Microscopy Images. *J. Chem. Inf. Model.* 2019, **60**:2492-2509 http://dx.doi.org/10.1021/acs.jcim.9b00734

[5] Raccuglia P, Elbert KC, Adler PDF, Falk C, Wenny MB, Mollo A, Zeller M, Friedler SA, Schrier J, Norquist AJ: Machine-Learning-Assisted Materials Discovery Using Failed Experiments. *Nature* 2016, **533**:73-76 http://dx.doi.org/10.1038/nature17439

[6] Jia X, Lynch A, Huang Y, Danielson M, Lang'at I, Milder A, Ruby AE, Wang H, Friedler SA, Norquist AJ, et al.: Anthropogenic Biases in Chemical Reaction Data Hinder Exploratory Inorganic Synthesis. *Nature* 2019, **573**:251-255 http://dx.doi.org/10.1038/s41586-019-1540-5

[7] Elena L. Cáceres, Nicholas C. Mew, and Michael J. Keiser, Adding Stochastic Negative Examples into Machine Learning Improves Molecular Bioactivity Prediction *Journal of Chemical Information and Modeling* **2020** 60 (12), 5957-5970 DOI: 10.1021/acs.jcim.0c00565

[8] Kirkpatrick, J., McMorrow, B., Turban, D. H., Gaunt, A. L., Spencer, J. S., Matthews, A. G., & Cohen, A. J. (2021). Pushing the frontiers of density functionals by solving the fractional electron problem. *Science*, *374*(6573), 1385-1389.

[9] Pernot, Pascal, and Andreas Savin. "Probabilistic performance estimators for computational chemistry methods: The empirical cumulative distribution function of absolute errors." *The Journal of Chemical Physics* 148, no. 24 (2018): 241707.

[10] McAnanama-Brereton, Suzanne, and Mark P. Waller. "Rational density functional selection using game theory." *Journal of Chemical Information and Modeling* 58, no. 1 (2018): 61-67.

[11] Kalinin, S. V., Ziatdinov, M., Spurgeon, S.R., Ophus, C., Stach, E.A., Susi, T., Agar, J., and Randall, J., 2022. Deep Learning for Electron and Scanning Probe Microscopy: From Materials Design to Atomic Fabrication. *MRS Bulletin*

[12] Unruh, D., Kolluru, V.S.C., Baskaran, A., Chen, Y. and Chan, M.K.Y., 2022. Theory+AI/ML for Microscopy and Spectroscopy – Challenges and Opportunities. *MRS Bulletin*

[13] Saar, L., Liang, H., Wang, A., McDannald A., Rodriguez, E., Takeuchi I. and Kusne A. G., 2022. The LEGOLAS Kit: A Low-Cost Robot Scientist for Education. *MRS Bulletin*
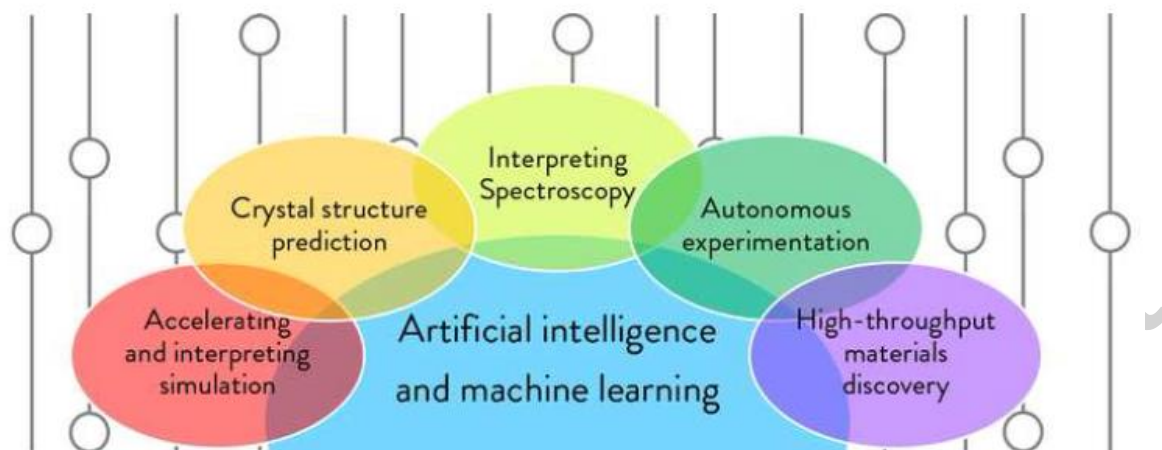
[14] Annevelink, E., Kurchin, R., Muckley, E., Kavalsky, L., Hegde, V., Sulzer, .V, Zhu, S., Pu, J., Farina, D., Johnson, M., Gandhi, D., Dave, A., Lin, H., Edelman, A., Ramsundar B., Saal, J., Rackauckas, C., Shah, V., Meredig, B. and Viswanathan, V. 2022., AutoMat: Automated Materials Discovery for Electrochemical systems. *MRS Bulletin*

[15] Yang, J. and Mannodi-Kanakkithodi, A., 2022. High-throughput Computations and Machine Learning for Halide Perovskite Discovery. *MRS Bulletin*

[16] Ceriotti, M., 2022. Beyond potentials: integrated machine-learning models for materials. *MRS Bulletin*

[17] Clements, R., Dickman, J., Johal, J., Martin, J., Glover, J. and Day, G.M., 2022. Roles and opportunities for machine learning in organic molecular crystal structure prediction and its applications. *MRS Bulletin*

[18] Sarupria, S., Hall, S.W. and Rogal, J., 2022. Machine learning for molecular simulations of crystal nucleation and growth. *MRS Bulletin*

[19] Liu, B., Qiu, Y., Goonetilleke E.C. and Huang, X., 2022. Kinetic Network Models to Study Molecular Self-assembly in the Wake of Machine Learning. *MRS Bulletin*

[20] Scheffler, M., Aeschlimann, M., Albrecht, M., Bereau, T., Bungartz, H.J., Felser, C., Greiner, M., Groß, A., Koch, C.T., Kremer, K. and Nagel, W.E., 2022. FAIR data enabling new horizons for materials research. *Nature*, *604*(7907), pp.635-642.

[21] Schmidt J, Marques MR, Botti S, Marques MA. Recent advances and applications of solid-state materials science. *npj Computational Materials*. 2019 Aug 8;5(1):1-36.

Pratyush Tiwary: Professor Pratyush Tiwary is a tenured Associate Professor at the University of Maryland, College Park in the Department of Chemistry and Biochemistry and the Institute for Physical Science and Technology. His work at the interface of molecular simulations, statistical mechanics and machine learning has been recognized through Sloan Research Fellowship in Chemistry, NSF CAREER award, NIH Maximizing Investigators' Research Award, ACS OpenEye Outstanding Junior Faculty Award and Doctoral New Investigator award from ACS-Petroleum Research Fund. He received his PhD in Materials Science from Caltech and undergraduate degree in Metallurgical Engineering from IIT-BHU, India. Email: ptiwary@umd.edu

Heather J. Kulik: Professor Heather J. Kulik is a tenured Associate Professor in the Department of Chemical Engineering at MIT. She received her B.E. in Chemical Engineering from the Cooper Union in 2004 and her Ph.D. from the Department of Materials Science and Engineering at MIT in 2009. From 2010-2013, she completed postdoctoral training at Lawrence Livermore and Stanford. Her research in machine learning and computational chemistry for transition metal catalysis has been recognized by numerous awards including the NSF CAREER and a Sloan Fellowship. Email: hjkulik@mit.edu