# Machine learning in materials research: developments over the last decade and challenges for the future

Anubhav Jain<sup>1\*</sup>

<sup>1</sup> Energy Technologies Area, Lawrence Berkeley National Laboratory, Berkeley CA 94720

\* Contact: ajain@lbl.gov

#### Abstract

The number of studies that apply machine learning (ML) to materials science has been growing at a rate of approximately 1.67 times per year over the past decade. In this review, I examine this growth in various contexts. First, I present an analysis of the most commonly used tools (software, databases, materials science methods, and ML methods) used within papers that apply ML to materials science. The analysis demonstrates that despite the growth of deep learning techniques, the use of classical machine learning is still dominant as a whole. It also demonstrates how new research can effectively build upon past research, particular in the domain of ML models trained on density functional theory calculation data. Next, I present the progression of best scores as a function of time on the matbench materials science benchmark for formation enthalpy prediction. In particular, a dramatic improvement of 7 times reduction in error is obtained when progressing from feature-based methods that use conventional ML (random forest, support vector regression, *etc.*) to the use of graph neural network techniques. Finally, I provide views on future challenges and opportunities, focusing on data size and complexity, extrapolation, interpretation, access, and relevance.

#### **1** Introduction

The use of machine learning techniques in materials research has grown in the last decade from a small niche topic to an entire subfield within materials science & engineering. Indeed, there have been over 2000 papers on the topic of materials machine learning in the year 2023 alone, and over the past decade there has been a 1.67 times yearly growth in the number of papers (Figure 1). A 2020 review by Morgan and Jacobs<sup>1</sup> found that not only were the number of papers on the topic exponentially increasing, but that the number of *review* papers per year on the topic had already reached nearly 40 by 2019. Indeed, there already exist many excellent reviews on various aspects of materials machine learning, including its applications in simulation and modeling<sup>2–4</sup>, synthesis and characterization<sup>5–7</sup>, manufacturing<sup>8,9</sup>, and literature mining<sup>10</sup>. Reviews also exist for specific topics such as structural materials<sup>11</sup> or best practices for research reporting<sup>12</sup>.



Figure 1 The number of publications per year on the topic of machine learning in materials science. Data was retrieved on Jan 23, 2024 via http://api.elsevier.com and http://www.scopus.com using the pybliometrics Python library. See Code and Data Availability for data collection code and extracted information.

This review both looks back and looks ahead. Looking back, it examines what has enabled the field of machine learning to advance so rapidly. Indeed, about five years ago it was unclear whether the field would enter a "trough of disillusionment" or an "AI winter"<sup>13</sup>. However, the development of the Transformer architecture<sup>14</sup> in the computer science domain and the crystal graph neural network<sup>15,16</sup> in the materials science domain around the same time revitalized much research and led to major advancements in performance. More recently, materials machine learning is rapidly building upon advances in natural language processing, and in particular large language models such as Generative Pretrained Transformer (GPT) models<sup>17</sup> (e.g., GPT-3, GPT-4, and ChatGPT). Thus, the field has largely avoided any periods of stagnation thus far and the pace of innovation appears to only be increasing.

One factor for the rapid growth of the field is the ability for materials machine learning research to rapidly build upon past work such as databases, software, ML methods, or domain-specific techniques. New research papers can build upon data sets from prior papers – bypassing expensive data collection and focusing on method development – or transfer methods developed by the computer science community to the materials domain. Such methods, packaged in reusable software libraries, can then be applied directly to tackling specific materials problems, often with minimal additional method development, data collection, or software programming.

This review is divided into three sections. In the first section, it presents an analysis of the cross-fertilization between machine learning methods, materials science methods, data and software by analyzing common citations between papers. The second section presents a

quantitative analysis of the amount of progress achieved in a particular subfield of materials machine learning – structure-based property prediction. The final section presents remaining challenges and opportunities, particularly in the areas of data size and complexity, extrapolation, interpretation, access, and relevance.



Figure 2 Progress in machine learning in materials science is stimulated by advances in software, materials databases, materials science methods, and machine learning methods.

#### 2 Rapid growth by building upon prior work

The rapid rise in publications on the topic of machine learning in materials science is fueled by advancements in software, material databases, domain-specific materials science methods, and domain-agnostic machine learning methods (Figure 2). To examine this relationship further beyond a simple publication count increase, I compiled a data set of 6795 research papers on the topic of materials machine learning and subsequently analyzed the citations for each of these papers using the Scopus API and pybliometrics<sup>15</sup> Python library. I then analyzed the data set to determine all papers that were cited at least 100 times *within* materials science machine learning papers to determine commonly used tools and techniques within this subdomain. We note that this method only counts citations *within* the set of 6795 materials machine learning papers, and therefore is lower than a full citation count which may include citations from many domains or study types.



#### Articles cited >100 times in materials science machine learning research

Figure 3 Analysis of papers with >100 citations within the collection of 6795 research papers collected on the topic of materials machine learning. Data was retrieved on Jan 23, 2024 via http://api.elsevier.com and http://www.scopus.com using the pybliometrics Python library. See Code and Data Availability for data collection code and extracted information.

First, I focust on the results for software (Figure 3a). The most commonly used software and the mostly highly cited work within the data set overall is the scikit-learn<sup>18</sup> Python package. This may stem from several reasons. First, scikit-learn implements a variety of techniques that can be useful for many different types of materials machine learning. For example, it can be applied to predict the band gaps of solids<sup>19</sup>, to predict the strength of cement composites<sup>20</sup>, to associate processing conditions with final properties in batteries<sup>21</sup>, to predict the fatigue life of powder metallurgy components<sup>22</sup>, or for many other materials tasks. Furthermore, although scikit-learn is missing the capability to implement more complex deep learning models, the small data set sizes of many materials problems often make it practical to use more conventional machine learning algorithms that have fewer parameters to train. Overall, it is interesting to note that the most commonly cited paper within the entire data set is a general-purpose machine learning tool rather than a domain-specific tool.

Following scikit-learn, the next most highly cited work is the VASP software (Figure 3a). Indeed, the next five software libraries (VASP<sup>23</sup>, pymatgen<sup>24</sup>, Phonopy<sup>25</sup>, matminer<sup>26</sup>, and AFLOW<sup>27</sup>) are generally used to calculate (VASP, Phonopy) or analyze (pymatgen, matminer, AFLOW) materials properties through density functional theory (DFT). The final software, OVITO<sup>28</sup>, is also used to visualize simulation results The high presence

these software libraries within the group of materials machine learning papers suggests that a large fraction of materials machine learning is being performed on simulation data. Analysis of top databases cited within MSE-ML papers (Figure 3b) supports this trend: all three databases (Materials Project<sup>29</sup>, OQMD<sup>30</sup>, and AFLOWLib.org<sup>31</sup>) focus primarily on density functional theory generated data sets.

When examining the most popular domain-agnostic machine learning methods that are applied to the materials domain (Figure 3c), traditional tree-based machine learning techniques are found to be the most dominant. The top two cited works are related to tree-based methods (random forest<sup>32</sup> and XGBoost<sup>33</sup>), and the third (Gradient Boosting<sup>34</sup>) is also frequently applied to tree-based methods. The use of deep learning<sup>35</sup> is becoming more popular; nevertheless, the larger data sets needed to train these algorithms likely inhibits more widespread usage. Finally, methods for interpreting machine learning methods, including SHAP model explanations<sup>36</sup> and t-SNE<sup>37</sup>, are also popular in the materials community.

Finally, I examine commonly used methods developed for the domain of materials science (Figure 3d). As with software and databases, simulation methods (DFT-GGA<sup>38</sup>, Molecular Dynamics<sup>39</sup>, DFT-Monkhorst Pack<sup>40</sup>, and Ab Initio Molecular Dynamics–AIMD<sup>41</sup>) make up much of the list. The next category of methods are descriptors for materials -i.e., methods that use crystal structure or composition as an input and produce a library of features that describe the input for use in machine learning techniques. Popular methods include Magpie descriptors<sup>42</sup>, SOAP descriptors<sup>43</sup>, Coloumb matrix<sup>44</sup>, moment tensor potentials<sup>45</sup>, Atom-centered symmetry<sup>46</sup>, and PL fragment descriptors<sup>47</sup>. A third category is methods for ML force fields, including Behler-Parrinello potentials<sup>48</sup>, GAP potentials<sup>49</sup>, SNAP potentials<sup>50</sup>, and Deep potentials<sup>51</sup>. Finally, two of the results are separate from the categories of simulation method, materials descriptor, or force field. The first is the crystal graph neural network<sup>15</sup> which popularized a neural network based approach to structureproperty relationships by solving the problem of crystal representation as a periodic graph. The second is Gaussian Processes for iterative exploration<sup>52</sup>, which is becoming an increasingly popular technique as machine learning is being integrated into for materials discovery campaigns.

Overall, what conclusions can we draw from such analysis? First, it is worth pointing out that despite the seeming dominance of deep learning techniques in achieving good performance in tasks like language modeling or image generation, in the materials domain such techniques are still overall less popular than conventional techniques. As found in the matbench study<sup>53</sup>, deep learning becomes much more attractive for larger data sets (a threshold of approximately 10,000 data points was found in that study). Many materials ML problems simply do not have the data to make effective use of deep learning techniques are now readily available for simulation-based machine learning. However, there still remains a need for similar large and concerted efforts in other domains of materials science such as synthesis, characterization, and materials processing data. Although such databases have been developed<sup>54–56</sup>, they do not factor heavily in our literature review of materials ML and are still in the process of realizing a large, coordinated research community. This may potentially be due to complexity and heterogeneity of data in experimental domains, making data preprocessing (e.g., cleaning

and standardization) cumbersome for subsequent ML. It is possible that the push for automation in experiments may help in this endeavor, however the complexity of describing materials samples in experiments makes this far from straightforward.

## **3** Steady gains in accuracy

The ability of materials machine learning to draw upon prior work has led to steady gains in accuracy for a variety of tasks. In particular, the degree of improvement can be quantified for the field of materials property prediction from crystal structures. As we will later show, the state-of-the-art-models today can outperform those from 7 years ago by over a factor of 7 improvement in accuracy. With the remarkable gains in accuracy comes the potential to do more science with greater confidence using machine learning models.

Task name	Samples	Algorithm	Verified MAE (unit) or ROC-AUC	Structure Required
matbench_steels	312	MODNet (v0.1.12)	87.763 (MPa)	
matbench_jdft2d	636	MODNet (v0.1.12)	33.192 (meV/atom)	
matbench_phonons	1265	MegNet (kgcnn v2.1.0)	28.761 (cm <sup>-1</sup> )	x
matbench_expt_gap	4604	MODNet (v0.1.12)	0.333 (eV)	
matbench_dielectric	4764	MODNet (v0.1.12)	0.271 (unitless)	
matbench_expt_is_metal	4921	AMMExpress v2020	0.921	
matbench_glass	5680	MODNet (v0.1.12)	0.960	
matbench_log_gvrh	10987	coNGN	0.067 (log <sub>10</sub> (GPa))	x
matbench_log_kvrh	10987	coNGN	0.049 (log <sub>10</sub> (GPa))	x
matbench_perovskites	18928	coGN	0.027 (eV/unit cell)	x
matbench_mp_gap	106113	coGN	0.156 (eV)	x
matbench_mp_is_metal	106113	CGCNN v2019	0.952	x
matbench_mp_e_form	132752	coGN	0.0170 (eV/atom)	x

Table 1 Current snapshot of matbench leaderboard (Jan 23, 2024).

To study the improvement of structure-property models in greater detail, I make use of the matbench<sup>53</sup> leaderboard (https://matbench.materialsproject.org). The leaderboard

measures the performance of various algorithms on a series of benchmark tasks. The current top performers for each of the 13 tasks is presented in Table 1. The tasks are ordered by the number of samples (*i.e.*, data points) in the task. There is a clear delineation in the leading algorithms based on the number of samples in the task. The MODNet algorithm<sup>57</sup> is the clear leader on tasks with <10,000 samples, leading 5 out of 7 tasks in this category. However, these particular tasks do not have a structure provided and only chemical composition is known. Meanwhile, for tasks with >10,000 samples, the (related) coNGN<sup>58</sup> and coGN<sup>58</sup> algorithms lead 5 out of 6 tasks. This separation based on sample size (with different leading algorithms at that time) was observed in the original matbench paper<sup>53</sup>. However, a more careful interpretation of the data shows that today, in all tasks where a structure is provided (even the phonons task with only 1,265 samples), some form of crystal graph neural network is the dominant algorithm. Algorithms such as MODNet<sup>57</sup> and AMMExpress<sup>53</sup> come into play only in tasks in which the structure is not provided, for which the crystal graph neural networks cannot be applied at all.



Figure 4 Progress made on a smaller data task (elastic constants) and a larger data task (formation enthalpy) by various machine learning models.

We next examine progress over time for one of the matbench tasks. In Figure 4, we plot the improvement in mean absolute error over time in predicting formation enthalpies of compounds as calculated by density functional theory simulations and tabulated in the Materials Project database. The most significant feature of Figure 4 is the sharp drop in mean absolute error when progressing from the random forest model with Magpie descriptors<sup>42</sup> to the crystal graph neural network<sup>15</sup> (CGCNN). The switch from more conventional machine learning techniques with a series of hand-tuned features to that of a neural network architecture with features that are largely learned on-the-fly led to a tremendous, immediate jump in performance (from >100 meV/atom to <35 meV/atom). Indeed, for large data tasks, it is generally the case that hand-tuned features are not necessary for good performance<sup>59</sup>. Subsequently, various extensions and improvements to the CGCNN architecture have steadily reduced the error further. State-the-art-models today

can reproduce formation enthalpies of compounds within the Materials Project database to <20 meV/atom, which is in most cases lower than the error of the density functional calculations as compared to experiments<sup>60</sup> and of various experiments amongst themselves<sup>61</sup>. At this point, the matbench task can likely be considered "solved" and more difficult tasks need to be designed.



Figure 5 Overview of major challenges and opportunities in materials ML. Current progress and future needs for each area are discussed in the main text.

#### 4 Future challenges and opportunities

Next, I briefly review progress and outline challenges in 5 areas that are outlined in Figure 5: data size and complexity, extrapolation, interpretation, access, and relevance.

**Data size and complexity**: Data is the essential raw material for machine learning. Unfortunately, materials data can be limited in quantity and high in complexity. Nevertheless, new advancements may help make progress in this area. The use of natural language processing techniques to parse the scientific literature has resulted in many new structured data sets being compiled from previous literature<sup>62</sup>. New advancements in large language models may allow researchers to extract data sets by simply providing a few examples of structured output from unstructured text<sup>63</sup>. Nevertheless, even though algorithmic improvements are rapid, gaining access to the raw literature data for parsing remains difficult. In parallel, researchers continue to compile data sets outside the materials modeling domain through user contributions<sup>54,55,64</sup>. From the analysis side, progress on small data remains more challenging as compared to large data sets. A previous study has generally found a power law scaling of performance with respect to data size for various graph neural network models<sup>65</sup>. Such scaling relies on larger data and is often ineffective

for small data sets. Therefore, techniques such as multi-fidelity modeling<sup>66,67</sup>, hybrid featurization and neural networks<sup>57</sup>, and transfer learning<sup>68</sup> may be needed to leverage big data sets when analyzing small data. While we are likely still far away from true "few shot" learning on general and extrapolative materials science tasks, progress has been reported in using large language models in this manner<sup>69,70</sup>. It is possible that foundational models, trained on large materials data sets and with general purpose "understanding" of materials science, will be able to serve as the basis for training more targeted machine learning algorithms using smaller data sets.

**Extrapolation**: No clear standard has emerged for how to quantify the extrapolative limits of machine learning models. Standard cross-validation is designed for interpolation; alternate methods such as LOCO-CV<sup>71</sup> are needed to ensure that the test set is significantly different than the training set for evaluating extrapolation. For example, several reports have indicated that the current set of crystal graph neural network algorithms, despite performing well on the matbench data set, may have issues generalizing out of distribution<sup>72,73</sup>. To this end, extensions like matbench-discovery<sup>74</sup> attempts to test on out of sample (*i.e.*, not in Materials Project<sup>29</sup>) structures to better test extrapolation. Nevertheless, the issue is far from solved. As training data sets grow in size and scope, it becomes more difficult to find and evaluate samples that are significantly outside the training domain. More fundamentally, evaluating ML models for open-ended exploration (*e.g.*, as in iterative machine learning or generative models) remains a fruitful area of research because standard performance metrics do not translate directly to these areas<sup>75,76</sup> and validating each new prediction for every algorithm can be expensive. Thus, many opportunties remain to better investigate the extrapolation capability of ML models.

**Interpretation:** Interpretable models are desirable because they may help uncover physical insights and relations rather than simply make predictions.<sup>77,78</sup> Interpretability can also clarify the domain under which the model is valid. Unfortunately, the general situation today is that the most accurate models also tend to be the most opaque. There exist many model-agnostic interpretation methods that can be applied on top of such models, such as partial dependence plots, individual conditional expectation plots, Shapley additive explanations<sup>36</sup>, or surrogate modeling. Nevertheless, these methods do not fully capture nor do they explain the decision-making of the underlying model. As an alternative, some recent work has focused on building interpretation within the model itself<sup>79</sup> or in restricting models to symbolic regression<sup>80</sup>. Unfortunately, without a clear metric for quantifying the interpretability of a model, it remains difficult to measure progress in this area.

**Access:** Access is increasingly becoming an issue; the popularity of proprietary models trained on proprietary data sets such as OpenAI's GPT have made it difficult to conduct open science using such models. Worse, such models can sometimes only be accessed by an API with the underlying performance and results of the "same" model changing over time<sup>81</sup>. Thus, researchers publishing a certain set of results at the beginning of a project may find that the results have changed when running the same analysis on ostensibly the same model towards the end of that project. Such behavior clearly poses issues for the reproducibility requirements of published scientific research. It is at present unclear what role such models have in the future of science. It is possible that studies incorporating such models are inadmissible outright, or it may be that the models may be considered as black

boxes (similar to human intuition or manual data processing) and can form part of the procedure so long as their output is independently verifiable by some other means.

The issue of access is not limited to large language models. For example, Google recently reported a breakthrough in the journal *Nature* for models to predict energies and forces from crystal structures<sup>82</sup>. However, neither the model nor the full data set required for training that model was made available, and attempting to reproducing the data would require exorbitant amounts of computing. As the resources needed to produce a state-of-the-art result increases, the issue of who will be able to access data and models becomes of greater concern.

**Relevance:** Finally, as the performance of machine learning models improves, it is worth keeping in mind that these models are intended to be a means to an end rather than the end in themselves. There are many examples where ML was helpful to an outcome even if it was not perfectly optimized<sup>83</sup>. Conversely, there are also examples where ML models obtain good scores on one metric that does not always translate to seemingly related metrics. Examples include formation enthalpy scores not translating to phase stability<sup>84</sup> or MAE scores not translating to materials discovery acceleration factors<sup>74</sup>. Furthermore, materials or design suggestions are not valuable unless they are eventually made and tested. To that end, machine learning in the "virtual" world needs to be more closely integrated with automated laboratories<sup>85,86</sup> and more attention should be paid to ways in which computational predictions might be validated.

## 5 Conclusion

The application of machine learning to materials research has seen remarkable transformations in a relatively short amount of time. Not too long ago, machine learning in materials was considered a niche field with relatively few publications. Today, with many thousands of articles being published yearly, keeping track of the various developments has become a major challenge. The citation analysis presented in this article provides clues as to the fuel behind the rapid advancements. New publications are built upon algorithms and tools not only from within the materials science community but also from the computer science community. This has led to rapid advancements in performance on benchmark tasks such as those in the matbench protocol.

Despite the challenges, outstanding challenges and questions still remain. In an age where model size and data set size translates directly to performance, what will be the role of individual research labs? The situation today is that few academic groups have the resources to reproduce state-of-the-art results from industry in many areas of machine learning, making the issue of access and reproducibility particularly concerning at the current moment. Furthermore, challenges related to data set size and complexity, extrapolation, interpretation, and relevance still require innovative solutions. Nevertheless, researchers continue to make progress in all areas, marking the current time as a particularly exciting era for machine learning in materials research.

## 6 Data and software availability

Data and analysis scripts for the literature analysis can be found at Github: https://github.com/computron/pybliometrics\_ml.

Data for the matbench analysis is derived from the archived benchmark data found in: https://github.com/materialsproject/matbench.

The raw data table used to derive the literature analysis plot is provided in the supplementary Excel file titled: "pybliometrics\_ml\_analysis.xlsx".

## 7 Acknowledgements

This work was primarily funded and intellectually led by the Materials Project, which is funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, under Contract no. DE-AC02-05-CH11231: Materials Project program KC23MP. The author would like to thank Dr. A. Dunn for assistance in the matbench analysis.

#### 8 References

- Morgan, D.; Jacobs, R. Opportunities and Challenges for Machine Learning in Materials Science. Annu. Rev. Mater. Res. 2020, 50 (1), 71–103. https://doi.org/10.1146/annurev-matsci-070218-010015.
- (2) Von Lilienfeld, O. A.; Burke, K. Retrospective on a Decade of Machine Learning for Chemical Discovery. *Nat Commun* **2020**, *11* (1), 4895. https://doi.org/10.1038/s41467-020-18556-9.
- (3) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine Learning for Molecular and Materials Science. *Nature* **2018**, *559* (7715), 547. https://doi.org/10.1038/s41586-018-0337-2.
- (4) Chen, C.; Zuo, Y.; Ye, W.; Li, X.; Deng, Z.; Ong, S. P. A Critical Review of Machine Learning of Energy Materials. *Advanced Energy Materials* 2020, *10* (8), 1903242. https://doi.org/10.1002/aenm.201903242.
- (5) Rickman, J. M.; Lookman, T.; Kalinin, S. V. Materials Informatics: From the Atomic-Level to the Continuum. *Acta Materialia* 2019, *168*, 473–510. https://doi.org/10.1016/j.actamat.2019.01.051.
- (6) Kalinin, S. V.; Ophus, C.; Voyles, P. M.; Erni, R.; Kepaptsoglou, D.; Grillo, V.; Lupini, A. R.; Oxley, M. P.; Schwenker, E.; Chan, M. K. Y.; Etheridge, J.; Li, X.; Han, G. G. D.; Ziatdinov, M.; Shibata, N.; Pennycook, S. J. Machine Learning in Scanning Transmission Electron Microscopy. *Nat Rev Methods Primers* **2022**, *2* (1), 1–28. https://doi.org/10.1038/s43586-022-00095-w.
- (7) Baskaran, A.; Kautz, E. J.; Chowdhary, A.; Ma, W.; Yener, B.; Lewis, D. J. Adoption of Image-Driven Machine Learning for Microstructure Characterization and Materials Design: A Perspective. *JOM* 2021, 73 (11), 3639–3657. https://doi.org/10.1007/s11837-021-04805-9.
- (8) Aggour, K. S.; Gupta, V. K.; Ruscitto, D.; Ajdelsztajn, L.; Bian, X.; Brosnan, K. H.; Chennimalai Kumar, N.; Dheeradhada, V.; Hanlon, T.; Iyer, N.; Karandikar, J.; Li, P.; Moitra, A.; Reimann, J.; Robinson, D. M.; Santamaria-Pang, A.; Shen, C.; Soare, M. A.; Sun, C.; Suzuki, A.; Venkataramana, R.; Vinciquerra, J. Artificial

Intelligence/Machine Learning in Manufacturing and Inspection: A GE Perspective. *MRS Bulletin* **2019**, *44* (7), 545–558. https://doi.org/10.1557/mrs.2019.157.

- (9) Bhuvaneswari, V.; Priyadharshini, M.; Deepa, C.; Balaji, D.; Rajeshkumar, L.; Ramesh, M. Deep Learning for Material Synthesis and Manufacturing Systems: A Review. *Materials Today: Proceedings* **2021**, *46*, 3263–3269. https://doi.org/10.1016/j.matpr.2020.11.351.
- (10) Olivetti, E. A.; Cole, J. M.; Kim, E.; Kononova, O.; Ceder, G.; Han, T. Y.-J.; Hiszpanski, A. M. Data-Driven Materials Research Enabled by Natural Language Processing and Information Extraction. *Applied Physics Reviews* **2020**, *7* (4), 041317. https://doi.org/10.1063/5.0021106.
- (11) Sparks, T. D.; Kauwe, S. K.; Parry, M. E.; Tehrani, A. M.; Brgoch, J. Machine Learning for Structural Materials. *Annu. Rev. Mater. Res.* 2020, 50 (1), 27–48. https://doi.org/10.1146/annurev-matsci-110519-094700.
- (12) Artrith, N.; Butler, K. T.; Coudert, F.-X.; Han, S.; Isayev, O.; Jain, A.; Walsh, A. Best Practices in Machine Learning for Chemistry. *Nat. Chem.* **2021**, *13* (6), 505–508. https://doi.org/10.1038/s41557-021-00716-z.
- (13) Reyes, K. G.; Maruyama, B. The Machine Learning Revolution in Materials? MRS Bull. 2019, 44 (7), 530–537. https://doi.org/10.1557/mrs.2019.153.
- (14) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In *Advances in Neural Information Processing Systems*; Curran Associates, Inc., 2017; Vol. 30.
- (15) Xie, T.; Grossman, J. C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys. Rev. Lett.* 2018, *120* (14), 145301. https://doi.org/10.1103/PhysRevLett.120.145301.
- (16) Reiser, P.; Neubert, M.; Eberhard, A.; Torresi, L.; Zhou, C.; Shao, C.; Metni, H.; Van Hoesel, C.; Schopmans, H.; Sommer, T.; Friederich, P. Graph Neural Networks for Materials Science and Chemistry. *Commun Mater* **2022**, *3* (1), 93. https://doi.org/10.1038/s43246-022-00315-6.
- (17) Brown, T. B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; Agarwal, S.; Herbert-Voss, A.; Krueger, G.; Henighan, T.; Child, R.; Ramesh, A.; Ziegler, D. M.; Wu, J.; Winter, C.; Hesse, C.; Chen, M.; Sigler, E.; Litwin, M.; Gray, S.; Chess, B.; Clark, J.; Berner, C.; McCandlish, S.; Radford, A.; Sutskever, I.; Amodei, D. Language Models Are Few-Shot Learners. arXiv July 22, 2020. https://doi.org/10.48550/arXiv.2005.14165.
- (18) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D. Scikit-Learn: Machine Learning in Python. *MACHINE LEARNING IN PYTHON*.
- (19) Kumar, U.; Mishra, K. A.; Kushwaha, A. K.; Cho, S. B. Bandgap Analysis of Transition-Metal Dichalcogenide and Oxide via Machine Learning Approach. *Journal of Physics and Chemistry of Solids* **2022**, *171*, 110973. https://doi.org/10.1016/j.jpcs.2022.110973.

- (20) Zhang, J.; Niu, W.; Yang, Y.; Hou, D.; Dong, B. Machine Learning Prediction Models for Compressive Strength of Calcined Sludge-Cement Composites. *Construction and Building Materials* 2022, 346, 128442. https://doi.org/10.1016/j.conbuildmat.2022.128442.
- (21) Turetskyy, A.; Wessel, J.; Herrmann, C.; Thiede, S. Battery Production Design Using Multi-Output Machine Learning Models. *Energy Storage Materials* 2021, *38*, 93– 112. https://doi.org/10.1016/j.ensm.2021.03.002.
- (22) Leininger, D. S.; Reissner, F.-C.; Baumgartner, J. New Approaches for a Reliable Fatigue Life Prediction of Powder Metallurgy Components Using Machine Learning. *Fatigue & Fracture of Engineering Materials & Structures* 2023, 46 (3), 1190–1210. https://doi.org/10.1111/ffe.13921.
- (23) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B* 1996, 54 (16), 11169– 11186. https://doi.org/10.1103/PhysRevB.54.11169.
- (24) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (Pymatgen): A Robust, Open-Source Python Library for Materials Analysis. *Computational Materials Science* 2013, 68, 314–319. https://doi.org/10.1016/j.commatsci.2012.10.028.
- (25) Togo, A.; Tanaka, I. First Principles Phonon Calculations in Materials Science. *Scripta Materialia* **2015**, *108*, 1–5. https://doi.org/10.1016/j.scriptamat.2015.07.021.
- (26) Ward, L.; Dunn, A.; Faghaninia, A.; Zimmermann, N. E. R.; Bajaj, S.; Wang, Q.; Montoya, J.; Chen, J.; Bystrom, K.; Dylla, M.; Chard, K.; Asta, M.; Persson, K. A.; Snyder, G. J.; Foster, I.; Jain, A. Matminer: An Open Source Toolkit for Materials Data Mining. *Computational Materials Science* **2018**, *152*, 60–69. https://doi.org/10.1016/j.commatsci.2018.05.018.
- (27) Curtarolo, S.; Setyawan, W.; Hart, G. L. W.; Jahnatek, M.; Chepulskii, R. V.; Taylor, R. H.; Wang, S.; Xue, J.; Yang, K.; Levy, O.; Mehl, M. J.; Stokes, H. T.; Demchenko, D. O.; Morgan, D. AFLOW: An Automatic Framework for High-Throughput Materials Discovery. *Computational Materials Science* 2012, 58, 218–226. https://doi.org/10.1016/j.commatsci.2012.02.005.
- (28) Stukowski, A. Visualization and Analysis of Atomistic Simulation Data with OVITO-the Open Visualization Tool. *Modelling Simul. Mater. Sci. Eng.* **2009**, *18* (1), 015012. https://doi.org/10.1088/0965-0393/18/1/015012.
- (29) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Materials* 2013, *1* (1), 011002. https://doi.org/10.1063/1.4812323.
- (30) Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* 2013, 65 (11), 1501–1509. https://doi.org/10.1007/s11837-013-0755-4.

- (31) Curtarolo, S.; Setyawan, W.; Wang, S.; Xue, J.; Yang, K.; Taylor, R. H.; Nelson, L. J.; Hart, G. L. W.; Sanvito, S.; Buongiorno-Nardelli, M.; Mingo, N.; Levy, O. AFLOWLIB.ORG: A Distributed Materials Properties Repository from High-Throughput Ab Initio Calculations. *Computational Materials Science* 2012, 58, 227–235. https://doi.org/10.1016/j.commatsci.2012.02.002.
- (32) Breiman, L. Random Forests. *Machine Learning* **2001**, *45* (1), 5–32. https://doi.org/10.1023/A:1010933404324.
- (33) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; KDD '16; Association for Computing Machinery: New York, NY, USA, 2016; pp 785–794. https://doi.org/10.1145/2939672.2939785.
- (34) Friedman, J. H. Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics* **2001**, *29* (5), 1189–1232.
- (35) LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, *521* (7553), 436–444. https://doi.org/10.1038/nature14539.
- (36) Lundberg, S. M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems; NIPS'17; Curran Associates Inc.: Red Hook, NY, USA, 2017; pp 4768– 4777.
- (37) Maaten, L. van der; Hinton, G. Visualizing Data Using T-SNE. *Journal of Machine Learning Research* 2008, 9 (86), 2579–2605.
- (38) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, 77 (18), 3865–3868. https://doi.org/10.1103/PhysRevLett.77.3865.
- (39) Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *Journal* of Computational Physics **1995**, 117 (1), 1–19. https://doi.org/10.1006/jcph.1995.1039.
- (40) Monkhorst, H. J.; Pack, J. D. Special Points for Brillouin-Zone Integrations. *Phys. Rev. B* 1976, 13 (12), 5188–5192. https://doi.org/10.1103/PhysRevB.13.5188.
- (41) Kresse, G.; Hafner, J. Ab Initio Molecular Dynamics for Liquid Metals. *Phys. Rev. B* 1993, 47 (1), 558–561. https://doi.org/10.1103/PhysRevB.47.558.
- (42) Ward, L.; Agrawal, A.; Choudhary, A.; Wolverton, C. A General-Purpose Machine Learning Framework for Predicting Properties of Inorganic Materials. *npj* Computational Materials 2016, 2 (1), 16028–16028. https://doi.org/10.1038/npjcompumats.2016.28.
- (43) Bartók, A. P.; Kondor, R.; Csányi, G. On Representing Chemical Environments. *Phys. Rev. B* 2013, 87 (18), 184115. https://doi.org/10.1103/PhysRevB.87.184115.
- (44) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* 2012, *108* (5), 058301. https://doi.org/10.1103/PhysRevLett.108.058301.

- (45) Shapeev, A. V. Moment Tensor Potentials: A Class of Systematically Improvable Interatomic Potentials. *Multiscale Model. Simul.* 2016, 14 (3), 1153–1173. https://doi.org/10.1137/15M1054183.
- (46) Behler, J. Atom-Centered Symmetry Functions for Constructing High-Dimensional Neural Network Potentials. *The Journal of Chemical Physics* 2011, *134* (7), 074106. https://doi.org/10.1063/1.3553717.
- (47) Isayev, O.; Oses, C.; Curtarolo, S.; Tropsha, A. Universal Fragment Descriptors for Predicting Electronic Properties of Inorganic Crystals. **2016**, 1–12.
- (48) Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* 2007, *98* (14), 146401. https://doi.org/10.1103/PhysRevLett.98.146401.
- (49) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Gaussian Approximation Potentials: The Accuracy of Quantum Mechanics, without the Electrons. *Physical Review Letters* 2010, 104 (April), 1–4. https://doi.org/10.1103/PhysRevLett.104.136403.
- (50) Thompson, A. P.; Swiler, L. P.; Trott, C. R.; Foiles, S. M.; Tucker, G. J. Spectral Neighbor Analysis Method for Automated Generation of Quantum-Accurate Interatomic Potentials. *Journal of Computational Physics* 2015, 285, 316–330. https://doi.org/10.1016/j.jcp.2014.12.018.
- (51) Zhang, L.; Han, J.; Wang, H.; Car, R.; E, W. Deep Potential Molecular Dynamics: A Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* 2018, *120* (14), 143001. https://doi.org/10.1103/PhysRevLett.120.143001.
- (52) Xue, D.; Balachandran, P. V.; Hogden, J.; Theiler, J.; Xue, D.; Lookman, T. Accelerated Search for Materials with Targeted Properties by Adaptive Design. *Nature Communications* 2016, 7, 11241–11241. https://doi.org/10.1038/ncomms11241.
- (53) Dunn, A.; Wang, Q.; Ganose, A.; Dopp, D.; Jain, A. Benchmarking Materials Property Prediction Methods: The Matbench Test Set and Automatminer Reference Algorithm. *npj Comput Mater* **2020**, *6* (1), 138. https://doi.org/10.1038/s41524-020-00406-3.
- (54) Puchala, B.; Tarcea, G.; Marquis, Emmanuelle. A.; Hedstrom, M.; Jagadish, H. V.; Allison, J. E. The Materials Commons: A Collaboration Platform and Information Repository for the Global Materials Community. *JOM* 2016, 68 (8), 2035–2044. https://doi.org/10.1007/s11837-016-1998-7.
- (55) Blaiszik, B.; Chard, K.; Pruyne, J.; Ananthakrishnan, R.; Tuecke, S.; Foster, I. The Materials Data Facility: Data Services to Advance Materials Science Research. *JOM* 2016, 68 (8), 2045–2052. https://doi.org/10.1007/s11837-016-2001-3.
- (56) Blokhin, E.; Villars, P. The PAULING FILE Project and Materials Platform for Data Science: From Big Data Toward Materials Genome. In *Handbook of Materials Modeling : Methods: Theory and Modeling*; Andreoni, W., Yip, S., Eds.; Springer International Publishing: Cham, 2018; pp 1–26. https://doi.org/10.1007/978-3-319-42913-7\_62-1.

- (57) De Breuck, P.-P.; Hautier, G.; Rignanese, G.-M. Materials Property Prediction for Limited Datasets Enabled by Feature Selection and Joint Learning with MODNet. *npj Comput Mater* 2021, 7 (1), 1–8. https://doi.org/10.1038/s41524-021-00552-2.
- (58) Ruff, R.; Reiser, P.; Stühmer, J.; Friederich, P. Connectivity Optimized Nested Graph Networks for Crystal Structures. arXiv August 9, 2023. https://doi.org/10.48550/arXiv.2302.14102.
- (59) Murdock, R. J.; Kauwe, S. K.; Wang, A. Y.-T.; Sparks, T. D. Is Domain Knowledge Necessary for Machine Learning Materials Properties? *Integr Mater Manuf Innov* 2020, 9 (3), 221–227. https://doi.org/10.1007/s40192-020-00179-z.
- (60) Hautier, G.; Ong, S. P.; Jain, A.; Moore, C. J.; Ceder, G. Accuracy of Density Functional Theory in Predicting Formation Energies of Ternary Oxides from Binary Oxides and Its Implication on Phase Stability. *Phys. Rev. B* 2012, 85 (15), 155208. https://doi.org/10.1103/PhysRevB.85.155208.
- (61) Kirklin, S.; Saal, J. E.; Meredig, B.; Thompson, A.; Doak, J. W.; Aykol, M.; Rühl, S.; Wolverton, C. The Open Quantum Materials Database (OQMD): Assessing the Accuracy of DFT Formation Energies. *npj Comput Mater* **2015**, *1* (1), 1–15. https://doi.org/10.1038/npjcompumats.2015.10.
- (62) Lee, J. H.; Lee, M.; Min, K. Natural Language Processing Techniques for Advancing Materials Discovery: A Short Review. *Int. J. of Precis. Eng. and Manuf.-Green Tech.* 2023, 10 (5), 1337–1349. https://doi.org/10.1007/s40684-023-00523-6.
- (63) Dagdelen, J.; Dunn, A.; Lee, S.; Walker, N.; Rosen, A. S.; Ceder, G.; Persson, K. A.; Jain, A. Structured Information Extraction from Scientific Text with Large Language Models. *Nat Commun* **2024**, *15* (1), 1418. https://doi.org/10.1038/s41467-024-45563x.
- (64) Huck, P.; Gunter, D.; Cholia, S.; Winston, D.; N'Diaye, A. T.; Persson, K. User Applications Driven by the Community Contribution Framework MPContribs in the Materials Project. *Concurrency and Computation: Practice and Experience* 2016, 28 (7), 1982–1993. https://doi.org/10.1002/cpe.3698.
- (65) Fung, V.; Zhang, J.; Juarez, E.; Sumpter, B. G. Benchmarking Graph Neural Networks for Materials Chemistry. *npj Comput Mater* **2021**, 7 (1), 84. https://doi.org/10.1038/s41524-021-00554-0.
- (66) Chen, C.; Zuo, Y.; Ye, W.; Li, X.; Ong, S. P. Learning Properties of Ordered and Disordered Materials from Multi-Fidelity Data. *Nat Comput Sci* 2021, 1 (1), 46–53. https://doi.org/10.1038/s43588-020-00002-x.
- (67) Annevelink, E.; Kurchin, R.; Muckley, E.; Kavalsky, L.; Hegde, V. I.; 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.; Viswanathan, V. AutoMat: Automated Materials Discovery for Electrochemical Systems. *MRS Bulletin* 2022, 47 (10), 1036–1044. https://doi.org/10.1557/s43577-022-00424-0.
- (68) Chen, C.; Ong, S. P. AtomSets as a Hierarchical Transfer Learning Framework for Small and Large Materials Datasets. *npj Comput Mater* **2021**, 7 (1), 173. https://doi.org/10.1038/s41524-021-00639-w.

- (69) Jablonka, K. M.; Schwaller, P.; Ortega-Guerrero, A.; Smit, B. Leveraging Large Language Models for Predictive Chemistry. ChemRxiv October 17, 2023. https://doi.org/10.26434/chemrxiv-2023-fw8n4-v3.
- (70) Maik Jablonka, K.; Ai, Q.; Al-Feghali, A.; Badhwar, S.; D. Bocarsly, J.; M. Bran, A.; Bringuier, S.; Catherine Brinson, L.; Choudhary, K.; Circi, D.; Cox, S.; Jong, W. A. de; L. Evans, M.; Gastellu, N.; Genzling, J.; Victoria Gil, M.; K. Gupta, A.; Hong, Z.; Imran, A.; Kruschwitz, S.; Labarre, A.; Lála, J.; Liu, T.; Ma, S.; Majumdar, S.; W. Merz, G.; Moitessier, N.; Moubarak, E.; Mouriño, B.; Pelkie, B.; Pieler, M.; Caldas Ramos, M.; Ranković, B.; G. Rodriques, S.; N. Sanders, J.; Schwaller, P.; Schwarting, M.; Shi, J.; Smit, B.; E. Smith, B.; Herck, J. V.; Völker, C.; Ward, L.; Warren, S.; Weiser, B.; Zhang, S.; Zhang, X.; Ahmad Zia, G.; Scourtas, A.; J. Schmidt, K.; Foster, I.; D. White, A.; Blaiszik, B. 14 Examples of How LLMs Can Transform Materials Science and Chemistry: A Reflection on a Large Language Model Hackathon. Digital Discovery 1233-1250. 2023. 2 (5), https://doi.org/10.1039/D3DD00113J.
- (71) Meredig, B.; Antono, E.; Church, C.; Hutchinson, M.; Ling, J.; Paradiso, S.; Blaiszik, B.; Foster, I.; Gibbons, B.; Hattrick-simpers, J.; Mehta, A. Can Machine Learning Identify the Next YBCO? Examining Extrapolation Performance for Materials Discovery Molecular Systems Design & Engineering. 2013, 1–3. https://doi.org/10.1039/x0xx00000x.
- (72) Li, K.; DeCost, B.; Choudhary, K.; Greenwood, M.; Hattrick-Simpers, J. A Critical Examination of Robustness and Generalizability of Machine Learning Prediction of Materials Properties. *npj Comput Mater* **2023**, *9* (1), 1–9. https://doi.org/10.1038/s41524-023-01012-9.
- (73) Omee, S. S.; Fu, N.; Dong, R.; Hu, M.; Hu, J. Structure-Based out-of-Distribution (OOD) Materials Property Prediction: A Benchmark Study. arXiv January 15, 2024. http://arxiv.org/abs/2401.08032 (accessed 2024-02-20).
- (74) Riebesell, J.; Goodall, R. E. A.; Benner, P.; Chiang, Y.; Deng, B.; Lee, A. A.; Jain, A.; Persson, K. A. Matbench Discovery -- A Framework to Evaluate Machine Learning Crystal Stability Predictions. arXiv February 4, 2024. https://doi.org/10.48550/arXiv.2308.14920.
- (75) Borg, C. K. H.; Muckley, E. S.; Nyby, C.; Saal, J. E.; Ward, L.; Mehta, A.; Meredig, B. Quantifying the Performance of Machine Learning Models in Materials Discovery. *Digital Discovery* 2023, 2 (2), 327–338. https://doi.org/10.1039/D2DD00113F.
- (76) Zhao, Y.; Siriwardane, E. M. D.; Wu, Z.; Fu, N.; Al-Fahdi, M.; Hu, M.; Hu, J. Physics Guided Deep Learning for Generative Design of Crystal Materials with Symmetry Constraints. *npj Comput Mater* **2023**, *9* (1), 1–12. https://doi.org/10.1038/s41524-023-00987-9.
- (77) Choubisa, H.; Todorović, P.; Pina, J. M.; Parmar, D. H.; Li, Z.; Voznyy, O.; Tamblyn, I.; Sargent, E. H. Interpretable Discovery of Semiconductors with Machine Learning. *npj Comput Mater* **2023**, *9* (1), 117. https://doi.org/10.1038/s41524-023-01066-9.

- (78) Esterhuizen, J. A.; Goldsmith, B. R.; Linic, S. Interpretable Machine Learning for Knowledge Generation in Heterogeneous Catalysis. *Nat Catal* **2022**, *5* (3), 175–184. https://doi.org/10.1038/s41929-022-00744-z.
- (79) Teufel, J.; Torresi, L.; Reiser, P.; Friederich, P. MEGAN: Multi-Explanation Graph Attention Network. In *Explainable Artificial Intelligence*; Longo, L., Ed.; Communications in Computer and Information Science; Springer Nature Switzerland: Cham, 2023; pp 338–360. https://doi.org/10.1007/978-3-031-44067-0\_18.
- (80) Sun, S.; Ouyang, R.; Zhang, B.; Zhang, T.-Y. Data-Driven Discovery of Formulas by Symbolic Regression. *MRS Bulletin* **2019**, *44* (7), 559–564. https://doi.org/10.1557/mrs.2019.156.
- (81) Chen, L.; Zaharia, M.; Zou, J. How Is ChatGPT's Behavior Changing over Time? arXiv October 31, 2023. https://doi.org/10.48550/arXiv.2307.09009.
- (82) Merchant, A.; Batzner, S.; Schoenholz, S. S.; Aykol, M.; Cheon, G.; Cubuk, E. D. Scaling Deep Learning for Materials Discovery. *Nature* **2023**, *624* (7990), 80–85. https://doi.org/10.1038/s41586-023-06735-9.
- (83) Saal, J. E.; Oliynyk, A. O.; Meredig, B. Machine Learning in Materials Discovery: Confirmed Predictions and Their Underlying Approaches. *Annu. Rev. Mater. Res.* 2020, 50 (1), annurev-matsci-090319-010954. https://doi.org/10.1146/annurevmatsci-090319-010954.
- (84) Bartel, C. J.; Trewartha, A.; Wang, Q.; Dunn, A.; Jain, A.; Ceder, G. A Critical Examination of Compound Stability Predictions from Machine-Learned Formation Energies. *npj Comput Mater* 2020, 6 (1), 97. https://doi.org/10.1038/s41524-020-00362-y.
- (85) Stach, E.; DeCost, B.; Kusne, A. G.; Hattrick-Simpers, J.; Brown, K. A.; Reyes, K. G.; Schrier, J.; Billinge, S.; Buonassisi, T.; Foster, I.; Gomes, C. P.; Gregoire, J. M.; Mehta, A.; Montoya, J.; Olivetti, E.; Park, C.; Rotenberg, E.; Saikin, S. K.; Smullin, S.; Stanev, V.; Maruyama, B. Autonomous Experimentation Systems for Materials Development: A Community Perspective. *Matter* 2021, *4* (9), 2702–2726. https://doi.org/10.1016/j.matt.2021.06.036.
- (86) Szymanski, N. J.; Rendy, B.; Fei, Y.; Kumar, R. E.; He, T.; Milsted, D.; McDermott, M. J.; Gallant, M.; Cubuk, E. D.; Merchant, A.; Kim, H.; Jain, A.; Bartel, C. J.; Persson, K.; Zeng, Y.; Ceder, G. An Autonomous Laboratory for the Accelerated Synthesis of Novel Materials. *Nature* 2023, 624 (7990), 86–91. https://doi.org/10.1038/s41586-023-06734-w.