Matter

CellPress

Perspective

Opportunities for machine learning to accelerate halide-perovskite commercialization and scale-up

Rishi E. Kumar,[1,2] Armi Tiihonen,[3] Shijing Sun,[3] David P. Fenning,[1,2,*] Zhe Liu,[3,4,*] and Tonio Buonassisi[3,*]

SUMMARY

While halide perovskites attract significant academic attention, examples of industrial production at scale are still sparse. In this perspective, we review practical challenges hindering the commercialization of halide perovskites and discuss how machine-learning (ML) tools could help: (1) active-learning algorithms that blend institutional knowledge and human expertise could help stabilize and rapidly update baseline manufacturing processes, (2) computer-imaging methods with ML-based classification tools could help narrow the performance gap between large- and small-area devices, and (3) inference methods could help accelerate root-cause analysis by reconciling multiple data streams and simulations, focusing research efforts on the highest-probability areas. We conclude that to tackle many of these challenges, incremental—not radical—adaptations of existing ML methods are needed. We propose how industry-academic partnerships could help adapt "ready-now" ML tools to specific industry needs, further improve process control by revealing underlying mechanisms, and develop "gamechanger" discovery-oriented algorithms to better navigate the vast spaces of materials choices.

INTRODUCTION

In under two decades, researchers propelled the AM1.5 efficiency of single-junction halide-perovskite photovoltaic (PV) devices beyond 25%,[1–4] in addition to demonstrating competitive performances in other optoelectronic applications including light-emitting diodes and photodetectors.[5,6] However, few successful examples exist of commercial production at scale.[7,8] Part of this is due to well-documented concerns surrounding perovskite instability; while impressive demonstrations of perovskite-module durability have been shown in Cheacharoen et al.,[9] Shi et al.,[10] and Bogachuk et al.[11] (i.e., passing the IEC damp-heat and humidity-freeze tests), perovskite stability remains a risk.[12,13] But conversations with industrial partners suggest that lesser-discussed and perhaps more mundane concerns exist, including, but not limited to, establishing a stable baseline process. This motivated us to engage in a broader and more open-ended conversation with industry worldwide about the outstanding challenges they face. We distilled these discussions into four overarching challenges that frame the body of this perspective on how applications of machine learning (ML) may help to address them (Figure 1).

Meanwhile, over the past decade, applied ML methods have started to gain increased industrial acceptance, aiding complex process optimization and diagnosis.[14–18] During informal discussions with the PV industry, it became apparent

Progress and Potential

Halide perovskites have excellent optoelectronic properties, which enable applications in photovoltaics, photodetectors, and light emitting diodes, among others. However, further commercialization of halide perovskites requires significant improvement in the Technology Readiness Level, which could typically take a decade or longer. Machine learning (ML) has recently emerged as a useful tool to solve the complex diagnosis and optimization problems. This perspective examines the opportunities for both the "ready-now" ML algorithms that directly tackle the practical challenges in the industry, and the future "gamechanger" machine intelligence that leads the way into new perovskite discovery. We also discuss ML adoption barriers, and potential pathways to overcome these barriers including closer industry-academic partnerships focused on tech transfer. Overall, we see this applied ML field as having great potential to accelerate halide perovskites development and deployment.

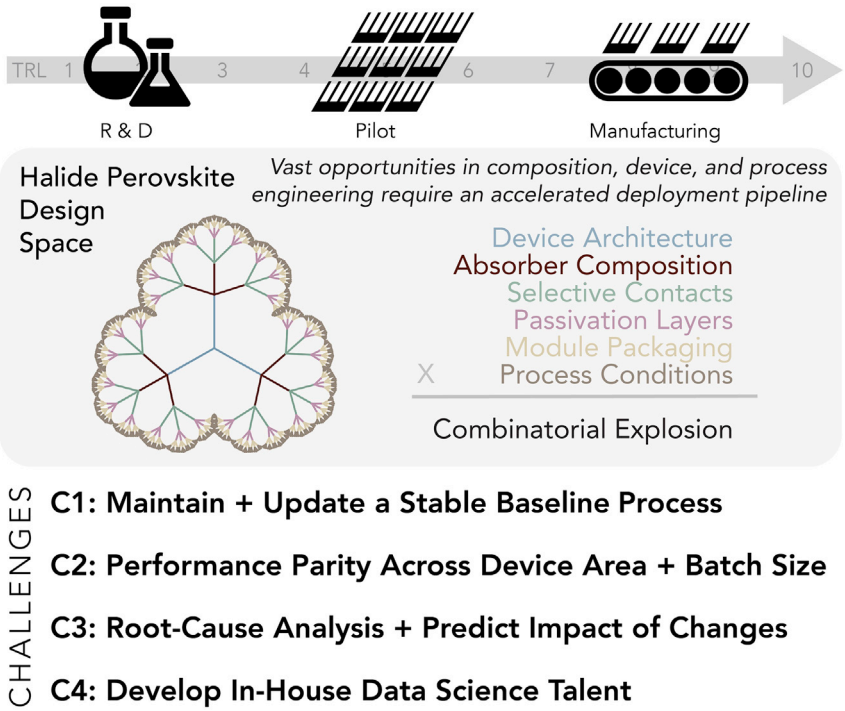

Figure 1. Challenges faced to manufacture perovskites at scale

Schematic of why R&D for scalable manufacturing technologies for perovskite becomes very complex, especially when increasing the technology readiness level (TRL). Generally, as the TRL initially increases, the halide-perovskite design space balloons. This manifests as several practical challenges, four of the most notable of which are listed here.

that ML methods are slow to gain acceptance, in part because of concerns surrounding interpretability, talent accessibility, and uncertainty about advantages over traditional designs of experiments.

These observations motivate the central questions of this perspective. (1) What are the key barriers that industry faces to commercialize halide-perovskite optoelectronics, and (2) can ML methods assist industry to overcome these challenges? We hope this perspective motivates future work, especially in the field of applied ML, aiding perovskites to "see the light of day."

WHAT CHALLENGES DOES INDUSTRY FACE TO MANUFACTURE PEROVSKITES AT SCALE?

Through discussions with industry worldwide, we observe a continuum of concerns spanning efforts across technology readiness levels (TRLs).[19] These concerns span from the "process-development phase" (developing a new process in research and development [R&D] and translating it to small lots in production, roughly TRLs 5–7) to the later "manufacturing scale-up phase" (achieving high performance in qualification runs, roughly TRLs 7–9). We distill and articulate industry needs in the following challenges:

C1. Maintaining a stable baseline process (as measured by efficiency and environmental stability) and frequently updating the process to incorporate the latest advances in the public domain

C2. Achieving large-area devices and/or large-scale manufacturing that perform close to small-area devices

[1]Department of Nanoengineering, UC San Diego, La Jolla, CA 92093, USA

[2]Materials Science & Engineering Program, UC San Diego, La Jolla, CA 92093, USA

[3]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[4]School of Materials Science & Engineering, Northwestern Polytechnical University, Xi'an, 710072 Shaanxi, P.R. China

*Correspondence: dfenning@eng.ucsd.edu (D.P.F.), zhe.liu@nwpu.edu.cn (Z.L.), buonassi@mit.edu (T.B.)

https://doi.org/10.1016/j.matt.2022.04.016

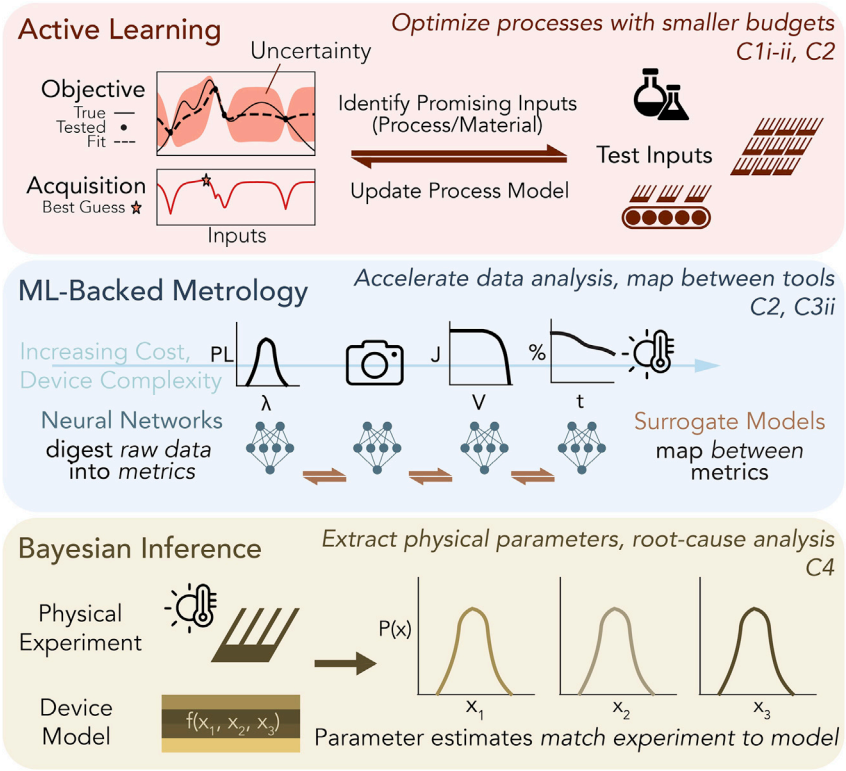

Figure 2. Existing machine-learning tools for perovskite commercialization

Schematic of how ready-now machine learning tools can assist perovskite-device commercialization. Techniques are labeled by the sections in which they are discussed within the text.

C3. Troubleshooting root causes of under-performance (ideally in-line and early in the manufacturing process) and estimating performance improvements arising from process or device-architecture modifications

C4. Developing in-house data-science talent

While most academic research to date on halide perovskites has focused on earlier TRLs, the translational activities that are required for perovskites to reach commercial readiness include substantial scientific unknowns regarding perovskite processing, where academic-industrial collaborations in new directions to directly address remaining challenges may prove fruitful. Furthermore, the academic research community has applied ML to other manufacturing systems[20–23] that directly address some of the needs above; successful examples of industrial ML scale-up include battery lifetime prediction for electric vehicles,[24,25] medical diagnostics,[26] and high-strength structural materials.[27,28] The next section explores opportunities to apply these learnings to perovskite commercialization.

HOW CAN ML METHODS ACCELERATE PEROVSKITE INDUSTRIALIZATION?

Here, we describe ML tools that may aid perovskite manufacturing and commercialization, using the aforementioned challenges as section headers. As industry places a premium in risk reduction, we emphasize existing ML-based approaches, with "existence proofs" in either perovskites or manufacturing in general (Figure 2). In the last section on "gamechangers," we describe how ongoing academic work applying

ML to perovskites connects to industry needs. For an introduction to ML methods applied to materials research, the reader is referred to recent reviews.[3,29–34]

Identifying and maintaining stable baselines that can be rapidly upgraded

It can be challenging to identify and maintain a baseline process that yields perovskite solar cells and modules with consistent performance and durability. Perovskites are sensitive to multiple spatially and temporally varying process parameters, including environmental conditions. The multinary nature of the bulk composition, as well as sensitivity to commonly uncontrolled variables (e.g., residual atmospheric solvent vapor pressure and relative humidity,[35] among others), can affect product quality. We consider how established ML methods may be able to assist in identifying and maintaining more stable baselines, especially when combined with stringent process control to minimize variance. Open-source development is becoming common practice in the field, and code ("background intellectual property [IP]") is often made available under licenses that grant users the right to implement and modify the open-source code in their own derivative work ("foreground IP"). All open-source licenses allow foreground IP to remain private, though some even allow foreground IP to be licensed or sold; "permissive" licenses (e.g., MIT, BSD) allow for the sublicensing, sale, and patenting of foreground IP, while "copy-left" licenses (e.g., GNU GPL) require that foreground IP be kept freely available.[36]

Find the baseline

Development of a baseline perovskite fabrication process can be a costly exercise in optimization across a high-dimensional design space consisting of various device layers and corresponding processing conditions. Active learning, including Bayesian optimization (BO),[37–40] presents an approach to identify optimal perovskite processing conditions with smaller experimental budgets. A recent pre-print by Liu et al. applied BO to optimize perovskite solar cell devices synthesized at Stanford by rapid-spray plasma processing (RSPP).[41] They demonstrate advantages over traditional design of experiments (DOE) approaches (including factorial sampling with partial-grid and one-variable-at-a-time searches), identifying multiple process conditions with higher performance after five batches of optimization. Sun et al.[42] used a related BO-based approach to identify a range of spin-coated perovskite absorber compositions with optimum environmental stability, using only 1.8% of available compositions. Active learning has also been demonstrated to be effective in guiding high-throughput computational screening of candidate perovskite materials.[43,44]

Active-learning methods have been applied to other device elements and related absorbers to identify baseline processing conditions that maximize conversion efficiency and reaction yields. BO has been used to guide experimental searches for fabrication conditions to maximize the conductivity of hole-transport materials.[45] Langner et al.[46] showed that BO of an organic PV (OPV) absorber layer defined in a quaternary compositional space takes 30× fewer samples than brute-force combinatorics. As a point of comparison, a similar reduction in the number of runs can be difficult to achieve in a high-resolution fractional factorial design (e.g., 1/32 fractionation) without significant *a priori* knowledge and confidence about which factors are suitable to confound. Several studies combined ML and robotics to autonomously fabricate and optimize perovskite single crystals, thin-film photoabsorbers, quantum dots, and charge-transport materials for perovskite solar devices.[45,47–50] BO has also been shown to be an effective search algorithm for the optimization of chemical-reaction yields, where the state-of-the-art yields for reactions defined in search spaces of 10^5 points were surpassed with experimental budgets of 10^1–10^2 points.[51] The

power of active-learning methods to reduce the number of experiments required to advance a figure of merit is increasingly well demonstrated and quite general.

Recent extensions to BO include multi-objective optimization to map Pareto fronts,[52–54] benchmarking work to optimize the degree of exploration and exploitation of the BO algorithm for experimental materials-science applications,[55–58] dimensionality-reduction approaches to address design spaces with >20 input variables,[46,51,59–62] and uncertainty quantification.[63–65] In summary, the literature has consistently shown that active-learning-driven optimization typically requires a total experimental budget of 10–100 unique sampling conditions—well within the usual throughput of an average team and often superior to human-driven experimental design.[45,46,51,66–68]

Maintain the baseline

Once a baseline process has been established, it must be tracked and maintained against external variations (e.g., changes in manufacturing ambient, chemical precursors, operators/tooling, etc.). Recent BO studies applied to manufacturing focus on identifying not only the highest performing but also the most robust process. This can be achieved by identifying relatively flat plateaus in the performance parameter of merit, e.g., using the open-source Golem algorithm.[69] To identify process drifts on the fly and predict ultimate performance, ML applied to luminescence imaging on semi-fabricates was used for silicon solar cells.[70,71] The predictive nature of ML models allows us to correlate the input synthesis and testing conditions with the device performance. When unexpected changes in the input conditions such as precursor contamination take place, in an ideal scenario, this shall be reflected in the reduction of predictive accuracy and serves as a starting point to review the baseline process for improved reproducibility. Periodic ML-driven sampling may nevertheless be needed to adjust baselines to slowly fluctuating uncontrolled variables, such as seasonal variation in ambient humidity that was demonstrated to affect perovskite crystal growth.[35] In fact, the adoption of ML-based process control provides greater flexibility to include both deliberate and circumstantial (e.g., environmental conditions, precursor batch) variables in process assessments, allowing one to catch unintended influences (like humidity) on the overall process performance. Note that dumping all possible input data into a model is a recipe for overfitting—the downselection of input features by their importance and redundance is a critical first step in ML model development.[59,72,73]

Upgrade the baseline

The target compositions and manufacturing processes of various device layers evolve rapidly, thanks to both internal and external R&D, but are typically challenging to incorporate into a baseline process. The ability to upgrade a baseline process is therefore of interest. Liu et al. report using prior machine settings as a "soft constraint" within a BO framework, allowing the search for the new optimum to leverage old information as a "warm start."[41] Sun et al.[42] and Gongora et al.[74] use first-principles simulations in a similar way.

Scale-up: Achieving large-area devices and/or large-scale manufacturing that perform close to small area

In the early days of perovskite development, increasing device area resulted in efficiency declines over twice as large as those in commercial inorganic PV materials.[7,75] Recent lab-scale demonstrations have reduced these losses,[76] although achieving high efficiencies with large active areas using manufacturing-friendly processes[77] remains a challenge.

Computer vision offers rapid identification of spatial and temporal inhomogeneities. When computer vision is combined with ML, the performance impacts of the aforementioned inhomogeneities can be predicted, and process-control corrections can be suggested.[78] Such metrology can be applied across TRLs, accumulating training data and building efficacy as they are moved from benchtop investigation to in-line-process monitoring. Well-established imaging-processing tools leverage the benefits of ML in handling large datasets of pixels and therefore provide an effective avenue in monitoring, quantifying, and controlling spatially varying parameters in large-area devices. Photoluminescence and electroluminescence images, coupled to device simulations and ML, enable in-line performance prediction for inorganic PV[79–82] and have recently been extended to perovskites.[65,83] ML algorithms developed by Tian et al.[84] and Taherimakhsousi et al.[85] estimate film thickness and defect density non-destructively by processing optical information. These ML-enabled methods augment (rather than substitute) existing efforts to tighten process controls, identify anomalies, identify uncontrolled variables, and reduce spatial and temporal non-uniformities, potentially enabling improved performance prediction earlier in the manufacturing line, early detection of process excursions and quality incidents, and more effective preventative maintenance.

Troubleshooting root causes of under-performance (ideally in-line and early in the manufacturing process) and estimating performance improvements arising from process or device-architecture modifications

Extract physical parameters of solar cells

The root causes of underperformance of solar cells can be extracted by pairing physics-based simulations with current-voltage-curve measurements. However, this method of troubleshooting can be time consuming and inconclusive, as the number of simulation fitting variables can be vast, and two or more limiting factors often combine to limit performance.

Bayesian inference is a probabilistic method to combine models and measurements in a statistically rigorous way, representing as probability distributions the underlying variables and their correlations. If one can provide a physical model to simulate the response of a system to experimental inputs, Bayesian inference allows the results of a physical experiment to be understood in the context of that model, enabling root-cause extraction from failure-analysis studies. Brandt et al.[86] and Kurchin et al.[87] demonstrated the application of Bayesian inference to underlying parameter extraction (e.g., minority-carrier lifetime, surface-recombination velocity, interfacial energy barriers, bulk-defect properties, etc.) for thin-film and silicon solar cells. Ren et al.[88] coupled Bayesian inference to a heuristic process model to design a time-temperature profile for a gallium arsenide (GaAs) solar cell device stack, improving device efficiency by 6.5% relative to a DOE best; essentially, the ML algorithm both optimized and identified the root causes of underperformance, informing the user how to improve the process. Oviedo et al. applied Bayesian inference to extract the root causes of degradation during environmental testing of OPV devices.[89] One of the advantages of this method is decoupling efficiency contributors in a finished device stack, where all device layers and interfaces are in their final state. A key enabler of these inference methods is that a neural network "surrogate" model can be trained to mimic a numerical device simulator over a range of input values, running 100–1,000× faster than the original simulator (see Figure S1 in Bommes et al.[82]). This approach enables even complex device simulators to be incorporated into Bayesian-inference algorithms, though care must be taken to ensure that the surrogate model's training data span the entire range of each inferred output with sufficient resolution such that the model is interpolating rather than extrapolating.

It may also be helpful to pair anomaly detection with such approaches to trigger a manual investigation when model uncertainty is high.

The challenges for wider adoption of Bayesian inference to emerging materials like perovskites are (1) to produce perovskite device-physics models that capture all performance-relevant underlying physics, including polarization, ion migration, and second-phase formation[90–93] and (2) to adapt Bayesian-inference models to high-efficiency devices with incremental improvements. To date, device efficiencies <20% have been studied, with large efficiency spreads between samples. In manufacturing, higher efficiencies and smaller spreads are more typical; the smaller signal to noise means both a more accurate model and higher-quality data are required, e.g., combining *J-V* measurements with other device measurements inside a Bayesian-inference framework.

Find early predictors for ultimate performance

ML models can help engineers establish correlations between characterization results within the fabrication or stability testing process as well as ultimate performance parameters of merit, saving time and resources. As an R&D example, optical images on bare perovskite films (which turn yellow during degradation) can give an indication of the environmental stability of finished perovskite devices;[94,95] this information can be combined with an active-learning algorithm to guide compositional tuning in perovskites, increasing device stability by ~2× within the ternary [Cs-MA-FA]-Pb-I compositional space with only a small fraction of this space fabricated into full devices.[42] For organic solar cells, Oviedo[89] developed a time-series forecast model that estimates device efficiency using only the first few hours (e.g., the initial 5%–10% of total testing time), significantly shortening the time required for degradation testing. For silicon solar cells, Kunze et al.[71] established an ML model to predict the final *I-V* metrics (i.e., J_{SC}, V_{OC}, *FF*, efficiency) based on luminescence images collected on the production line. Stoddard et al.[96] and Howard et al.[97] applied recurrent neural networks to predict the photoluminescence dynamics of lead-halide perovskite thin films under humidity cycling, demonstrating the value of ML in performing effective time-series forecasting for perovskite optoelectronic behaviors. Despite these encouraging examples, we are currently at an early stage in predicting the operational reliability of perovskite solar cells, close collaborations between device fabrication and long-term performance testing and the seamless integration between physical and data-driven models will be critical to develop effective prediction capabilities.

Developing in-house data-science talent

In the late 1990s, when website design required coding, a high premium was paid to hire rare HTML talent. As graphical user interfaces (GUIs) evolved throughout the 2000s, basic website-making became accessible to anyone with a computer, while professionals specialized in higher-end services. ML is currently at a similar early stage in its technology adoption cycle, with many companies and open-source communities launching software-as-a-service products. As ML code libraries, and GUI wrappers around ML libraries, become more commonplace and a dominant design emerges, it is foreseeable that basic ML capabilities will become as accessible as word processors are today.

Navigating this early phase of ML adoption can be challenging for perovskite manufacturers, and it can be tempting just to "wait it out"[98] But not upskilling existing staff in ML methods carries not only opportunity cost but also timeline risk because unfamiliarity breeds mistrust, which delays ML adoption. Here are some resources for the busy employee and manager to upskill in applied ML:

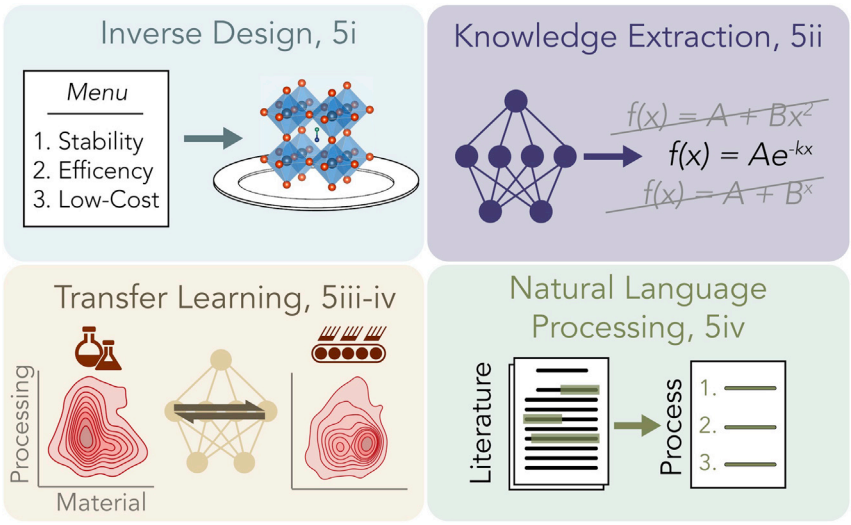

Figure 3. Upcoming machine-learning tools for perovskite commercialization
Game-changing ML techniques for photovoltaic research, development, and deployment. The sections in which these methods are described are labeled in each bubble.

- Applied ML tutorials and repositories of practical examples are increasingly available. Examples include MLOps[99] and Accelerated Materials Development for Manufacturing.[100]
- The above examples often include open-source code on repositories like GitHub or Papers with Code. These codes and libraries can often be adapted to your problem. Stack Overflow is a most useful community for troubleshooting.
- To keep abreast of literature, we prefer digests that access arXiv pre-prints (which are becoming common practice in applied ML) such as Google Scholar. Social media is also a surprisingly effective distribution channel for pre-prints; Twitter has an active computational chemistry (#compchem) community, and a LinkedIn group by Benji Maruyama on Autonomous Research Systems[101] filters many relevant papers. Peer review has new embodiments when code and datasets can be locally tested within minutes.

The tools described here represent our perspective as of 2021-Q4. This is a fast-evolving field. We are likely in the "fluid" phase described by Utterback,[102] where a variety of ML technology options exist prior to convergence on a dominant design. In this phase, adaptiveness is essential. For example, one us recently upgraded our group's BO capabilities from GPyOpt (development and maintenance ended in 2020[103]) to BoTorch,[104] which required a quarter-long, full-time-employee investment.

Gamechangers

Ongoing academic research into applied ML, if successful, could meaningfully aid perovskite commercialization, providing new and potentially transformative tools. These "gamechanger" techniques (Figure 3) are described below in no particular order.

Inverse design tools capable of solving real-world problems

Successful inverse design capabilities could enable the discovery of new perovskite composition spaces that achieve improved environmental stability, sustained high performance over a wider range of illumination conditions (e.g., for tandems or non-AM1.5 operating conditions), removal of lead, improved resilience to temporal and spatial variance in manufacturing, or some combination thereof. Most

experiments and simulation tools use materials and process variables as inputs and calculate "properties" as outputs. Solving the "inverse problem" involves designing an algorithm that instead uses desired properties as inputs to generate a prospective material and process—even suggesting materials not included in existing perovskite databases[105,106] or density functional theory (DFT) calculations.[107] While this vision was formulated for some time,[108,109] the rise of variational autoencoders (VAEs) enabled researchers to encode property-process-structure relationships within a generative model.[110–113] Related approaches include using generative adversarial networks (GANs) or genetic algorithms (GAs) to generate candidate compounds through directed evolution. Outstanding challenges include filtering out candidate materials that cannot be synthesized,[59,114–117] recommending synthesis pathways for those that can,[112,118] and coupling inverse-design algorithms to materials-optimization platforms to experimentally test new candidates.

Extracting physical insights from ML models

Successful "knowledge extraction"—derivation of greater physical insight—from ML models could enable both greater acceptance of ML methods by human researchers and greater generalization of results. This topic constitutes the ML subfields of explainable artificial intelligence and scientific ML.[119–121] A straightforward method uses ML to "fit" unknown input parameters of simulations,[122–124] with similar advantages and disadvantages as Bayesian inference. Knowledge can also be extracted from trained ML algorithms in the form of feature-importance ranking, e.g., Shapley Additive Explanations (SHAP[125,126]) and Local Interpretable Model-agnostic Explanations (LIME[127]), which have been used to illuminate how input materials and process parameters affect perovskite environmental stability;[49,94] however, they must be applied with caution to avoid user bias.[128,129] ML methods that construct decision trees[130] are inherently interpretable, exemplified by Kong et al.'s[131] reduction of materials databases into design rules to determine inorganic crystal structures. Scientific ML is an emergent class of models with the objective of finding an equation (often from a library of plausible functional forms) that best describes the relationship between data inputs and outputs. Naik et al.[132] used scientific ML to determine a rate equation for the degradation of $MAPbI_3$ films, the form of which suggested that film degradation proceeds by an autocatalytic reaction.

Facilitating process transfer between sites, equipment, and operators

The replication of perovskite syntheses across even operators within the same laboratory has proven notoriously difficult. If a synthesis or manufacturing process can be captured via an ML model, transfer learning may provide a means to rapidly tune the process to a new setting. If successful, transfer learning could accelerate recipe adaptation between manufacturing lines or within the same manufacturing line over time (longitudinally), the transfer of new processes from R&D into manufacturing, and possibly between manufacturing methods with reasonably similar input and output parameters (e.g., between blade coating and slot-die coating). While the raw input-process variables will differ, the embedded information contained within an accurately trained ML model for perovskite process optimization should contain an embedding of the underlying physics and chemistry (e.g., crystal-growth kinetics and thermodynamics) critical to desired outputs, like device performance. Successful embedding of the high-dimensional manufacturing process may be either data driven or possibly designed by using physics-informed neural networks. But, ultimately, the alignment of the critical information embedded inside two ML models from different processes (i.e., transfer learning) may still be plausible when the input and/or output spaces vary. Such "heterogeneous" transfer

learning has been demonstrated in other domains for tasks such as cross-lingual natural-language processing or process optimization for injection molding of different parts.[133–135] Successful transfer learning across solar cell manufacturing processes would be game changing, but its feasibility remains to be shown.

Automatically digesting the literature into the research and development pipeline

New materials and processes are constantly being described in the perovskite literature. Advances in natural-language processing (NLP) could facilitate programmatic extraction and deployment of freshly published processes into the R&D pipeline, augmenting all previously mentioned tools presented in this article. While such a platform remains hypothetical for perovskite research, Mehr et al.[136] have demonstrated a nearly autonomous workflow for robotic execution of chemical syntheses mined from the literature. The NLP toolset required to distill the literature is ever growing,[118,137–140] with text-mining tools like ChemDataExtractor[141] and ChemicalTagger[142] and image-mining tools like EXSCLAIM![143] (Extraction, Separation, and Caption-based natural Language Annotation of Images) currently available for download. New candidate materials or recipes could be explored using a combination of voting rules, transfer learning, and active learning to rapidly test high-probability candidates and automatically incorporate innovations into manufacturing lines.

CONCLUSIONS

The exquisite sensitivity of perovskites to local-composition and process variables, the vastness of possibilities for single layers and layer combinations, the rapid expansion of the academic literature, and the pressing need to deliver technologies for sustainability encourage the adoption of an R&D framework that arbitrages the significant advances made in ML. ML algorithms hold promise to rapidly and adaptively downselect and optimize complex design spaces for techno-economically relevant outputs. We present a summary of "ready-now" ML algorithms that may aid industrial perovskite-device development. We posit that in many cases, ML adoption is now an incremental advancement, within reach of many industry teams, with code and datasets freely available online. To help "bridge the gap," we advocate for greater interaction between academic and industry teams concerning the topic of applied ML.

ACKNOWLEDGMENTS

The authors thank their industrial contacts, group members, and collaborators for numerous helpful discussions that shaped this perspective.

AUTHOR CONTRIBUTIONS

Conceptualization, all authors; investigation, R.E.K., D.P.F., Z.L., and T.B.; writing – original draft, R.E.K., D.P.F., Z.L., and T.B.; writing – review & editing, all authors; supervision, D.P.F., Z.L., and T.B.

DECLARATION OF INTERESTS

Z.L. and T.B. own equity in a startup, Xinterra, focused on applying ML to develop novel materials.

REFERENCES

1. Li, D., Zhang, D., Lim, K.-S., Hu, Y., Rong, Y., Mei, A., Park, N.-G., and Han, H. (2021). A review on scaling up perovskite solar cells. Adv. Funct. Mater. *31*, 2008621. https://doi.org/10.1002/adfm.202008621.

2. Perini, C.A.R., Doherty, T.A.S., Stranks, S.D., Correa-Baena, J.-P., and Hoye, R.L.Z. (2021). Pressing challenges in halide perovskite photovoltaics—from the atomic to module level. Joule *5*, 1024–1030. https://doi.org/10.1016/j.joule.2021.03.011.

3. Tao, Q., Xu, P., Li, M., and Lu, W. (2021). Machine learning for perovskite materials design and discovery. NPJ Comput. Mater. *7*, 1–18. https://doi.org/10.1038/s41524-021-00495-8.

4. NREL Champion photovoltaic module efficiency chart (Rev. 05/10/2021). https://www.nrel.gov/pv/assets/pdfs/champion-module-efficiencies.pdf.

5. Veldhuis, S.A., Boix, P.P., Yantara, N., Li, M., Sum, T.C., Mathews, N., and Mhaisalkar, S.G. (2016). Perovskite materials for light-emitting diodes and lasers. Adv. Mater. *28*, 6804–6834. https://doi.org/10.1002/adma.201600669.

6. Wu, H., Ge, Y., Niu, G., and Tang, J. (2021). Metal halide perovskites for X-ray detection and imaging. Matter *4*, 144–163. https://doi.org/10.1016/j.matt.2020.11.015.

7. Extance, A. (2019). The reality behind solar power's next star material. Nature *570*, 429–432. https://doi.org/10.1038/D41586-019-01985-Y.

8. Crownhart, C. (2021). Can the most exciting new solar material live up to its hype?. https://www.technologyreview.com/2021/06/29/1027451/perovskite-solar-panels-hype-commercial-debut/.

9. Cheacharoen, R., Boyd, C.C., Burkhard, G.F., Leijtens, T., Raiford, J.A., Bush, K.A., Bent, S.F., and McGehee, M.D. (2018). Encapsulating perovskite solar cells to withstand damp heat and thermal cycling. Sustainable Energy Fuels *2*, 2398–2406. https://doi.org/10.1039/C8SE00250A.

10. Shi, L., Bucknall, M.P., Young, T.L., Zhang, M., Hu, L., Bing, J., Lee, D.S., Kim, J., Wu, T., Takamure, N., et al. (2020). Gas chromatography mass spectrometry analyses of encapsulated stable perovskite solar cells. Science *368*, eaba2412. https://doi.org/10.1126/science.aba2412.

11. Bogachuk, D., Saddedine, K., Martineau, D., Narbey, S., Verma, A., Gebhardt, P., Herterich, J.P., Glissmann, N., Zouhair, S., Markert, J., et al. (2021). Perovskite photovoltaic devices with carbon-based electrodes withstanding reverse-bias voltages up to –9 V and surpassing IEC 61215:2016 International Standard. Solar RRL, 2100527. https://doi.org/10.1002/solr.202100527.

12. Khenkin, M.V., Katz, E.A., Abate, A., Bardizza, G., Berry, J.J., Brabec, C., Brunetti, F., Bulović, V., Burlingame, Q., Di Carlo, A., et al. (2020). Consensus statement for stability assessment and reporting for perovskite photovoltaics based on ISOS procedures. Nat. Energy *5*, 35–49. https://doi.org/10.1038/s41560-019-0529-5.

13. Dunfield, S.P., Bliss, L., Zhang, F., Luther, J.M., Zhu, K., van Hest, M.F.A.M., Reese, M.O., and Berry, J.J. (2020). From defects to degradation: a mechanistic understanding of degradation in perovskite solar cell devices and modules. Adv. Energy Mater. *10*, 1904054. https://doi.org/10.1002/aenm.201904054.

14. Carvalho, T.P., Soares, F.A.A.M.N., Vita, R., Francisco, R.d.P., Basto, J.P., and Alcalá, S.G.S. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. Comput. Ind. Eng. *137*, 106024. https://doi.org/10.1016/j.cie.2019.106024.

15. Weichert, D., Link, P., Stoll, A., Rüping, S., Ihlenfeldt, S., and Wrobel, S. (2019). A review of machine learning for the optimization of production processes. Int. J. Adv. Manuf. Technol. *104*, 1889–1902. https://doi.org/10.1007/s00170-019-03988-5.

16. Savage, N. (2021). Tapping into the Drug Discovery Potential of AI (Biopharma Dealmakers).

17. Abràmoff, M.D., Lavin, P.T., Birch, M., Shah, N., and Folk, J.C. (2018). Pivotal trial of an autonomous AI-based diagnostic system for detection of diabetic retinopathy in primary care offices. NPJ Digital Med. *1*, 39. https://doi.org/10.1038/s41746-018-0040-6.

18. Patel, H., Prajapati, D., Mahida, D., and Shah, M. (2020). Transforming petroleum downstream sector through big data: a holistic review. J. Pet. Explor. Prod. Technol. *10*, 2601–2611. https://doi.org/10.1007/s13202-020-00889-2.

19. Mankins, J. (1995). Technology Readiness Level – A White Paper (Office of Space Access and Technology, NASA). www.researchgate.net/publication/247705707_Technology_Readiness_Level_-_A_White_Paper.

20. Mäkinen, S., Skogström, H., Laaksonen, E., and Mikkonen, T. (2021). Who needs MLOps: what data scientists seek to accomplish and how can MLOps help?.

21. Miranda, L. (2021). Towards data-centric machine learning: a short review. jvmiranda921.github.io/notebook/2021/07/30/data-centric-ml/.

22. Wuest, T., Irgens, C., and Thoben, K.-D. (2014). An approach to quality monitoring in manufacturing using supervised machine learning on product state data. J. Intell. Manuf. *25*, 1167–1180. https://doi.org/10.1007/s10845-013-0761-y.

23. Monostori, L., Márkus, A., and Brussel, H. (1996). Machine learning approaches to manufacturing. CIRP Ann. Manuf. Technol. *45*, 675–712. https://doi.org/10.1016/S0007-8506(18)30216-6.

24. Aykol, M., Herring, P., and Anapolsky, A. (2020). Machine learning for continuous innovation in battery technologies. Nat. Rev. Mater. *5*, 725–727. https://doi.org/10.1038/s41578-020-0216-y.

25. Severson, K.A., Attia, P.M., Jin, N., Perkins, N., Jiang, B., Yang, Z., Chen, M.H., Aykol, M., Herring, P.K., Fraggedakis, D., et al. (2019). Data-driven prediction of battery cycle life before capacity degradation. Nat. Energy *4*, 383–391. https://doi.org/10.1038/s41560-019-0356-8.

26. Topol, E.J. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nat. Med. *25*, 44–56. https://doi.org/10.1038/s41591-018-0300-7.

27. Zhuang, H. (2021). From evidence to new high-entropy alloys. Nat. Comput. Sci. *1*, 458–459. https://doi.org/10.1038/s43588-021-00100-4.

28. Martin, J.H., Yahata, B.D., Hundley, J.M., Mayer, J.A., Schaedler, T.A., and Pollock, T.M. (2017). 3D printing of high-strength aluminium alloys. Nature *549*, 365–369. https://doi.org/10.1038/nature23894.

29. Li, J., Lim, K., Yang, H., Ren, Z., Raghavan, S., Chen, P.-Y., Buonassisi, T., and Wang, X. (2020). AI applications through the whole life cycle of material discovery. Matter *3*, 393–432. https://doi.org/10.1016/j.matt.2020.06.011.

30. Butler, K.T., Davies, D.W., Cartwright, H., Isayev, O., and Walsh, A. (2018). Machine learning for molecular and materials science. Nature *559*, 547–555. https://doi.org/10.1038/s41586-018-0337-2.

31. Correa-Baena, J.-P., Hippalgaonkar, K., van Duren, J., Jaffer, S., Chandrasekhar, V.R., Stevanovic, V., Wadia, C., Guha, S., and Buonassisi, T. (2018). Accelerating materials development via automation, machine learning, and high-performance computing. Joule *2*, 1410–1420. https://doi.org/10.1016/j.joule.2018.05.009.

32. Tabor, D.P., Roch, L.M., Saikin, S.K., Kreisbeck, C., Sheberla, D., Montoya, J.H., Dwaraknath, S., Aykol, M., Ortiz, C., Tribukait, H., et al. (2018). Accelerating the discovery of materials for clean energy in the era of smart automation. Nat. Rev. Mater. *3*, 5–20. https://doi.org/10.1038/s41578-018-0005-z.

33. Stach, E., DeCost, B., Kusne, A.G., Hattrick-Simpers, J., Brown, K.A., Reyes, K.G., Schrier, J., Billinge, S., Buonassisi, T., Foster, I., et al. (2021). Autonomous experimentation systems for materials development: a community perspective. Matter *4*, 2702–2726. https://doi.org/10.1016/j.matt.2021.06.036.

34. Ahmadi, M., Ziatdinov, M., Zhou, Y., Lass, E.A., and Kalinin, S.V. (2021). Machine learning for high-throughput experimental exploration of metal halide perovskites. Joule *5*, 2797–2822. https://doi.org/10.1016/j.joule.2021.10.001.

35. Nega, P.W., Li, Z., Ghosh, V., Thapa, J., Sun, S., Hartono, N.T.P., Nellikkal, M.A.N., Norquist, A.J., Buonassisi, T., Chan, E.M., and Schrier, J. (2021). Using automated serendipity to discover how trace water promotes and inhibits lead halide perovskite crystal formation. Appl. Phys. Lett. *119*, 041903. https://doi.org/10.1063/5.0059767.

36. Opensource.org. The Open Source Initiative. https://opensource.org/.

37. Deringer, V.L., Bartók, A.P., Bernstein, N., Wilkins, D.M., Ceriotti, M., and Csányi, G. (2021). Gaussian process regression for materials and molecules. Chem. Rev. *121*, 10073–10141. https://doi.org/10.1021/acs.chemrev.1c00022.

38. Chang, J., Nikolaev, P., Carpena-Núñez, J., Rao, R., Decker, K., Islam, A.E., Kim, J., Pitt, M.A., Myung, J.I., and Maruyama, B. (2020). Efficient closed-loop maximization of carbon nanotube growth rate using bayesian optimization. Sci. Rep. *10*, 9040. https://doi.org/10.1038/s41598-020-64397-3.

39. Zhang, Y., Apley, D.W., and Chen, W. (2020). Bayesian optimization for materials design with mixed quantitative and qualitative variables. Sci. Rep. *10*, 4924. https://doi.org/10.1038/s41598-020-60652-9.

40. Häse, F., Aldeghi, M., Hickman, R.J., Roch, L.M., and Aspuru-Guzik, A. (2021). Gryffin: an

algorithm for Bayesian optimization of categorical variables informed by expert knowledge. Appl. Phys. Rev. *8*, 031406. https://doi.org/10.1063/5.0048164.

41. Liu, Z., Rolston, N., Flick, A.C., Colburn, T.W., Ren, Z., Dauskardt, R.H., et al. (2022). Machine learning with knowledge constraints for process optimization of open-air perovskite solar cell manufacturing. Joule *6*, 834–849. https://doi.org/10.1016/j.joule.2022.03.003.

42. Sun, S., Tiihonen, A., Oviedo, F., Liu, Z., Thapa, J., Zhao, Y., Hartono, N.T.P., Goyal, A., Heumueller, T., Batali, C., et al. (2021). A data fusion approach to optimize compositional stability of halide perovskites. Matter *4*, 1305–1322. https://doi.org/10.1016/j.matt.2021.01.008.

43. Herbol, H.C., Hu, W., Frazier, P., Clancy, P., and Poloczek, M. (2018). Efficient search of compositional space for hybrid organic–inorganic perovskites via Bayesian optimization. NPJ Comput. Mater. *4*, 1–7. https://doi.org/10.1038/s41524-018-0106-7.

44. Chen, X., Wang, C., Li, Z., Hou, Z., and Yin, W.-J. (2020). Bayesian optimization based on a unified figure of merit for accelerated materials screening: a case study of halide perovskites. Sci. China Mater. *63*, 1024–1035. https://doi.org/10.1007/s40843-019-1255-4.

45. MacLeod, B.P., Parlane, F.G.L., and Morrissey, T.D. (2020). Self-driving laboratory for accelerated discovery of thin-film materials. Sci. Adv. *6*, eaaz8867.

46. Langner, S., Hase, F., Perea, J.D., Stubhan, T., Hauch, J., Roch, L.M., Heumueller, T., Aspuru-Guzik, A., and Brabec, C.J. (2020). Beyond ternary OPV: high-throughput experimentation and self-driving laboratories optimize multicomponent systems. Adv. Mater. *32*, e1907801. https://doi.org/10.1002/adma.201907801.

47. Epps, R.W., Bowen, M.S., Volk, A.A., Abdel-Latif, K., Han, S., Reyes, K.G., Amassian, A., and Abolhasani, M. (2020). Artificial chemist: an autonomous quantum dot synthesis bot. Adv. Mater. *32*, 2001626. https://doi.org/10.1002/adma.202001626.

48. Li, Z., Najeeb, M.A., Alves, L., Sherman, A.Z., Shekar, V., Cruz Parrilla, P., Pendleton, I.M., Wang, W., Nega, P.W., Zeller, M., et al. (2020). Robot-accelerated perovskite investigation and discovery. Chem. Mater. *32*, 5650–5663. https://doi.org/10.1021/acs.chemmater.0c01153.

49. Zhao, Y., Zhang, J., Xu, Z., Sun, S., Langner, S., Hartono, N.T.P., Heumueller, T., Hou, Y., Elia, J., Li, N., et al. (2021). Discovery of temperature-induced stability reversal in perovskites using high-throughput robotic learning. Nat. Commun. *12*, 2191. https://doi.org/10.1038/s41467-021-22472-x.

50. Kirman, J., Johnston, A., Kuntz, D.A., Askerka, M., Gao, Y., Todorović, P., Ma, D., Privé, G.G., and Sargent, E.H. (2020). Machine-learning-accelerated perovskite crystallization. Matter *2*, 938–947. https://doi.org/10.1016/j.matt.2020.02.012.

51. Shields, B.J., Stevens, J., Li, J., Parasram, M., Damani, F., Alvarado, J.I.M., Janey, J.M., Adams, R.P., and Doyle, A.G. (2021). Bayesian reaction optimization as a tool for chemical synthesis. Nature *590*, 89–96. https://doi.org/10.1038/s41586-021-03213-y.

52. MacLeod, B.P., Parlane, F.G.L., Dettelbach, K.E., Elliott, M.S., Rupnow, C.C., Morrissey, T.D., Haley, T.H., Proskurin, O., Rooney, M.B., Taherimakhsousi, N., et al. (2022). Advancing the Pareto Front Using a Self-Driving Laboratory. Nat. Commun. *13*, 995. https://doi.org/10.1038/s41467-022-28580-6.

53. Erps, T., Foshey, M., Luković, M.K., Shou, W., Goetzke, H.H., Dietsch, H., Stoll, K., Vacano, B.v., and Matusik, W. (2021). Accelerated Discovery of 3D Printing Materials Using Data-Driven Multi-Objective Optimization. Science Advances *7*, eabf7435. https://doi.org/10.1126/sciadv.abf7435.

54. Schweidtmann, A.M., Clayton, A.D., Holmes, N., Bradford, E., Bourne, R.A., and Lapkin, A.A. (2018). Machine learning meets continuous flow chemistry: automated optimization towards the Pareto front of multiple objectives. Chem. Eng. J. *352*, 277–282. https://doi.org/10.1016/j.cej.2018.07.031.

55. Rohr, B., Stein, H.S., Guevarra, D., Wang, Y., Haber, J.A., Aykol, M., Suram, S.K., and Gregoire, J.M. (2020). Benchmarking the acceleration of materials discovery by sequential learning. Chem. Sci. *11*, 2696–2706. https://doi.org/10.1039/C9SC05999G.

56. Häse, F., Aldeghi, M., Hickman, R.J., Roch, L.M., Christensen, M., Liles, E., Hein, J.E., and Aspuru-Guzik, A. (2021). Olympus: a benchmarking framework for noisy optimization and experiment planning. Mach. Learn. Sci. Technol. *2*, 035021. https://doi.org/10.1088/2632-2153/abedc8.

57. Liang, Q., Gongora, A.E., Ren, Z., Tiihonen, A., Liu, Z., Sun, S., Deneault, J.R., Bash, D., Mekki-Berrada, F., Khan, S.A., et al. (2021). Benchmarking the Performance of Bayesian Optimization across Multiple Experimental Materials Science Domains. NPJ Comput. Mater *7*, 188. https://doi.org/10.1038/s41524-021-00656-9.

58. Graff, D.E., Shakhnovich, E.I., and Coley, C.W. (2021). Accelerating high-throughput virtual screening through molecular pool-based active learning. Chem. Sci. *12*, 7866–7881. https://doi.org/10.1039/D0SC06805E.

59. Ye, W., Chen, C., Wang, Z., Chu, I.H., and Ong, S.P. (2018). Deep neural networks for accurate predictions of crystal stability. Nat. Commun. *9*, 3800. https://doi.org/10.1038/s41467-018-06322-x.

60. Tiihonen, A., Cox-Vazquez, S.J., Liang, Q., Ragab, M., Ren, Z., Hartono, N.T.P., Liu, Z., Sun, S., Zhou, C., Incandela, N.C., et al. (2021). Predicting Antimicrobial Activity of Conjugated Oligoelectrolyte Molecules via Machine Learning.

61. Wang, Z., Gehring, C., Kohli, P., and Jegelk, S. (2018). Batched large-scale Bayesian optimization in high-dimensional spaces. In International Conference on Artificial Intelligence and Statistics (PMLR), pp. 745–754. https://proceedings.mlr.press/v84/wang18c.html.

62. Lim, Y.-F., Ng, C.K., Vaitesswar, U.S., and Hippalgaonkar, K. (2021). Extrapolative bayesian optimization with Gaussian process and neural network ensemble surrogate models. Adv. Intell. Syst. *3*, 2100101. https://doi.org/10.1002/aisy.202100101.

63. Abdar, M., Pourpanah, F., Hussain, S., Rezazadegan, D., Liu, L., Ghavamzadeh, M., Fieguth, P., Cao, X., Khosravi, A., Acharya, U.R., et al. (2021). A Review of Uncertainty Quantification in Deep Learning: Techniques, Applications and Challenges. Information Fusion *76*, 243–297. https://doi.org/10.1016/j.inffus.2021.05.008.

64. Pfingsten, T. (2006). Bayesian Active Learning for Sensitivity Analysis. European Conference on Machine Learning (Springer).

65. Hirschfeld, L., Swanson, K., Yang, K., Barzilay, R., and Coley, C.W. (2020). Uncertainty quantification using neural networks for molecular property prediction. J. Chem. Inf. Model. *60*, 3770–3780. https://doi.org/10.1021/acs.jcim.0c00502.

66. Burger, B., Maffettone, P.M., Gusev, V.V., Aitchison, C.M., Bai, Y., Wang, X., Li, X., Alston, B.M., Li, B., Clowes, R., et al. (2020). A mobile robotic chemist. Nature *583*, 237–241. https://doi.org/10.1038/s41586-020-2442-2.

67. Yuan, R., Liu, Z., Balachandran, P.V., Xue, D., Zhou, Y., Ding, X., Sun, J., Xue, D., and Lookman, T. (2018). Accelerated discovery of large electrostrains in $BaTiO_3$ -based piezoelectrics using active learning. Adv. Mater. *30*, 1702884. https://doi.org/10.1002/adma.201702884.

68. Solomou, A., Zhao, G., Boluki, S., Joy, J.K., Qian, X., Karaman, I., Arróyave, R., and Lagoudas, D.C. (2018). Multi-objective Bayesian materials discovery: application on the discovery of precipitation strengthened NiTi shape memory alloys through micromechanical modeling. Mater. Des. *160*, 810–827. https://doi.org/10.1016/j.matdes.2018.10.014.

69. Aldeghi, M., Häse, F., Hickman, R.J., Tamblyn, I., and Aspuru-Guzik, A. (2021). Golem: an algorithm for robust experiment and process optimization. Preprint at arXiv. https://doi.org/10.48550/arXiv.2103.03716.

70. Demant, M., Virtue, P., Kovvali, A., Yu, S.X., and Rein, S. (2019). Learning quality rating of as-cut mc-Si wafers via convolutional regression networks. IEEE J. Photovolt. *9*, 1064–1072. https://doi.org/10.1109/JPHOTOV.2019.2906036.

71. Kunze, P., Rein, S., Hemsendorf, M., Ramspeck, K., and Demant, M. (2021). Learning an empirical digital twin from measurement images for a comprehensive quality inspection of solar cells. Solar RRL *Online Version*, 2100483, *in press*. https://doi.org/10.1002/solr.202100483.

72. Venkatesh, B.A.,J. (2019). A review of feature selection and its methods. Cybernetics Inf. Tech. *19*, 3–26.

73. Chandrashekar, G., and Sahin, F. (2014). A survey on feature selection methods. Comput. Electr. Eng. *40*, 16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024.

74. Gongora, A.E., Snapp, K.L., Whiting, E., Riley, P., Reyes, K.G., Morgan, E.F., and Brown, K.A. (2021). Using simulation to accelerate autonomous experimentation: a case study using mechanics. iScience *24*, 102262. https://doi.org/10.1016/j.isci.2021.102262.

75. Li, Z., Klein, T.R., Kim, D.H., Yang, M., Berry, J.J., van Hest, M.F.A.M., and Zhu, K. (2018). Scalable fabrication of perovskite solar cells. Nat. Rev. Mater. *3*, 18017. https://doi.org/10.1038/natrevmats.2018.17.

76. Yang, Y., Xue, Z., Chen, L., Lau, C.F.J., and Wang, Z. (2021). Large-area perovskite films for PV applications: a perspective from nucleation and crystallization. J. Energy Chem. *59*, 626–641. https://doi.org/10.1016/j.jechem.2020.12.001.

77. Hoang, M.T., Ünlü, F., Martens, W., Bell, J., Mathur, S., and Wang, H. (2021). Towards the environmentally friendly solution processing of metal halide perovskite technology. Green. Chem. *23*, 5302–5336. https://doi.org/10.1039/D1GC01756J.

78. Siemenn, A.E., Shaulsky, E., Beveridge, M., Buonassisia, T., Hashmi, S.M., and Droric, I. (2021). Autonomous Optimization of Fluid Systems at Varying Length Scales. Preprint at arXiv *2105*, 13553.

79. XSolar-Hetero | photovoltaic simulation platform. https://www.xsolar-hetero.sg/.

80. Anand, G., Ke, C., Wong, J., Aberle, A., and Stangl, R. (2016). An Online, Web Based Solar Cell Simulation Interface for the Personalized Simulation of Various Solar Cell Architectures, Using Various Simulation Programs (EU PVSEC). https://doi.org/10.13140/RG.2.1.3289.7521.

81. Zhao, Y., Zhan, K., Wang, Z., and Shen, W. (2021). Deep learning-based automatic detection of multitype defects in photovoltaic modules and application in real production line. Prog. Photovolt. Res. Appl. *29*, 471–484. https://doi.org/10.1002/pip.3395.

82. Bommes, L., Pickel, T., Buerhop-Lutz, C., Hauch, J., Brabec, C., and Peters, I.M. (2021). Computer vision tool for detection, mapping, and fault classification of photovoltaics modules in aerial IR videos. Prog. Photovolt. Res. Appl. https://doi.org/10.1002/pip.3448.

83. Buratti, Y., Abdullah-Vetter, Z., Sowmya, A., Trupke, T., and Hameiri, Z. (2021). A Deep Learning Approach for Loss-Analysis from Luminescence Images. 20-25 June 2021 (IEEE), pp. 0097–0100. 0097-0100. 10.1109/PVSC43889.2021.9518512.

84. Tian, S.I.P., Liu, Z., Chellappan, V., Lim, Y.F., Ren, Z., Oviedo, F., Teo, B.H., Thapa, J., Dutta, R., MacLeod, B.P., et al. (2020). Rapid and accurate thin film thickness extraction via UV-Vis and machine learning. In 47th IEEE Photovoltaic Specialists Conference (PVSC).

85. Taherimakhsousi, N., MacLeod, B.P., Parlane, F.G.L., Morrissey, T.D., Booker, E.P., Dettelbach, K.E., and Berlinguette, C.P. (2020). Quantifying defects in thin films using machine vision. NPJ Comput. Mater. *6*, 1–6. https://doi.org/10.1038/s41524-020-00380-w.

86. Brandt, R.E., Kurchin, R.C., Steinmann, V., Kitchaev, D., Roat, C., Levcenco, S., Ceder, G., Unold, T., and Buonassisi, T. (2017). Rapid photovoltaic device characterization through bayesian parameter estimation. Joule *1*, 843–856. https://doi.org/10.1016/j.joule.2017.10.001.

87. Kurchin, R.C., Poindexter, J.R., Vähänissi, V., Savin, H., Cañizo, C.d., and Buonassisi, T. (2020). How much physics is in a current–voltage curve? Inferring defect properties from photovoltaic device measurements. IEEE J. Photovolt. *10*, 1532–1537. https://doi.org/10.1109/JPHOTOV.2020.3010105.

88. Ren, Z., Oviedo, F., Thway, M., Tian, S.I.P., Wang, Y., Xue, H., Dario Perea, J., Layurova, M., Heumueller, T., Birgersson, E., et al. (2020). Embedding physics domain knowledge into a Bayesian network enables layer-by-layer process innovation for photovoltaics. NPJ Comput. Mater. *6*, 1–9. https://doi.org/10.1038/s41524-020-0277-x.

89. Oviedo, F. (2020). Accelerated Development of Photovoltaics by Physics-Informed Machine Learning. Ph.D (Massachusetts Institute of Technology).

90. Le Corre, V.M., Stolterfoht, M., Perdigón Toro, L., Feuerstein, M., Wolff, C., Gil-Escrig, L., Bolink, H.J., Neher, D., and Koster, L.J.A. (2019). Charge transport layers limiting the efficiency of perovskite solar cells: how to optimize conductivity, doping, and thickness. ACS Appl. Energy Mater. *2*, 6280–6287. https://doi.org/10.1021/acsaem.9b00856.

91. Le Corre, V.M., Wang, Z., Koster, L.J.A., and Tress, W. (2020). Device modeling of perovskite solar cells: insights and outlooks. In Soft-Matter Thin Film Solar Cells, pp. 1–32. https://doi.org/10.1063/9780735422414_004.

92. Lopez-Varo, P., Jiménez-Tejada, J.A., García-Rosell, M., Ravishankar, S., Garcia-Belmonte, G., Bisquert, J., and Almora, O. (2018). Device physics of hybrid perovskite solar cells: theory and experiment. Adv. Energy Mater. *8*, 1702772. https://doi.org/10.1002/aenm.201702772.

93. Tessler, N., and Vaynzof, Y. (2020). Insights from device modeling of perovskite solar cells. ACS Energy Lett. *5*, 1260–1270. https://doi.org/10.1021/acsenergylett.0c00172.

94. Hartono, N.T.P., Thapa, J., Tiihonen, A., Oviedo, F., Batali, C., Yoo, J.J., Liu, Z., Li, R., Marrón, D.F., Bawendi, M.G., et al. (2020). How machine learning can help select capping layers to suppress perovskite degradation. Nat. Commun. *11*, 4172. https://doi.org/10.1038/s41467-020-17945-4.

95. Hashmi, S.G., Tiihonen, A., Martineau, D., Ozkan, M., Vivo, P., Kaunisto, K., Ulla, V., Zakeeruddin, S.M., and Grätzel, M. (2017). Long term stability of air processed inkjet infiltrated carbon-based printed perovskite solar cells under intense ultra-violet light soaking. J. Mater. Chem. A *5*, 4797–4802. https://doi.org/10.1039/C6TA10605F.

96. Stoddard, R.J., Dunlap-Shohl, W.A., Qiao, H., Meng, Y., Kau, W.F., and Hillhouse, H.W. (2020). Forecasting the decay of hybrid perovskite performance using optical transmittance or reflected dark-field imaging. ACS Energy Lett. *5*, 946–954. https://doi.org/10.1021/acsenergylett.0c00164.

97. Howard, J.M., Wang, Q., Lee, E., Lahoti, R., Gong, T., Srivastava, M., Abate, A., and Leite, M.S. (2020). Quantitative predictions of photo-emission dynamics in metal halide perovskites via machine learning. Preprint at arXiv. https://doi.org/10.48550/arXiv.2010.03702.

98. Gottbrath, C., Bailin, J., Meakin, C., Thompson, T., and Charfman, J.J. (1999). The Effects of Moore's Law and Slacking on Large Computations. Preprint at arXiv (arXiv). astro-ph/9912202.

99. Machine Learning Engineering for Production (MLOps) Specialization. https://www.coursera.org/specializations/machine-learning-engineering-for-production-mlops.

100. Accelerated Materials Development. www.youtube.com/channel/UCxaokYYzFI9XPOUP_W_sD9g

101. LinkedIn Group. Autonomous research systems. www.linkedin.com/groups/12176428/.

102. Utterback, J.M. (1994). Mastering the Dynamics of Innovation (Harvard Business Review Press).

103. GitHub - GPyOpt. github.com/SheffieldML/GPyOpt/blob/master/README.md.

104. Balandat, M., Karrer, B., Jiang, D.R., Daulton, S., Letham, B., Wilson, A.G., and Bakshy, E. (2020). BoTorch: A Framework for Efficient Monte-Carlo Bayesian Optimization. Advances in neural information processing systems *33*, 21524–21538.

105. Marchenko, E.I., Fateev, S.A., Petrov, A.A., Korolev, V.V., Mitrofanov, A., Petrov, A.V., Goodilin, E.A., and Tarasov, A.B. (2020). Database of two-dimensional hybrid perovskite materials: open-access collection of crystal structures, band gaps, and atomic partial charges predicted by machine learning. Chem. Mater. *32*, 7383–7388. https://doi.org/10.1021/acs.chemmater.0c02290.

106. Cai, Y., Xie, W., Teng, Y.T., Harikesh, P.C., Ghosh, B., Huck, P., Persson, K.A., Mathews, N., Mhaisalkar, S.G., Sherburne, M., and Asta, M. (2019). High-throughput computational study of halide double perovskite inorganic compounds. Chem. Mater. *31*, 5392–5401. https://doi.org/10.1021/acs.chemmater.9b00116.

107. Kim, C., Pilania, G., and Ramprasad, R. (2016). Machine learning assisted predictions of intrinsic dielectric breakdown strength of ABX3 perovskites. J. Phys. Chem. C *120*, 14575–14580. https://doi.org/10.1021/acs.jpcc.6b05068.

108. Zunger, A. (2018). Inverse design in search of materials with target functionalities. Nat. Rev. Chem. *2*, 0121. https://doi.org/10.1038/s41570-018-0121.

109. Franceschetti, A., and Zunger, A. (1999). The inverse band-structure problem of finding an atomic configuration with given electronic properties. Nature *402*, 60–63. https://doi.org/10.1038/46995.

110. Gómez-Bombarelli, R., Wei, J.N., Duvenaud, D., Hernández-Lobato, J.M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T.D., Adams, R.P., and Aspuru-Guzik, A. (2018). Automatic chemical design using a data-driven continuous representation of molecules. ACS Cent. Sci. *4*, 268–276. https://doi.org/10.1021/acscentsci.7b00572.

111. Ren, Z., Noh, J., Tian, S., Oviedo, F., Xing, G., Liang, Q., Aberle, A., Liu, Y., Li, Q., Jayavelu, S., et al. (2022). An invertible crystallographic representation for general inverse design of inorganic crystals with targeted properties*5* (Matter), pp. 314–335. https://doi.org/10.1016/j.matt.2021.11.032.

112. Griffiths, R.-R., and Hernández-Lobato, J.M. (2020). Constrained Bayesian optimization for automatic chemical design using variational autoencoders. Chem. Sci. *11*, 577–586. https://doi.org/10.1039/C9SC04026A.

113. Sattari, K., Xie, Y., and Lin, J. (2021). Data-driven algorithms for inverse design of polymers. Soft Matter *17*, 7607–7622. https://doi.org/10.1039/d1sm00725d.

114. McDermott, M.J., Dwaraknath, S.S., and Persson, K.A. (2021). A graph-based network for predicting chemical reaction pathways in solid-state materials synthesis. Nat. Commun. *12*, 3097. https://doi.org/10.1038/s41467-021-23339-x.

115. Vasylenko, A., Gamon, J., Duff, B.B., Gusev, V.V., Daniels, L.M., Zanella, M., Shin, J.F., Sharp, P.M., Morscher, A., Chen, R., et al. (2021). Element selection for crystalline inorganic solid discovery guided by unsupervised machine learning of experimentally explored chemistry. Nat. Commun. *12*, 5561. https://doi.org/10.1038/s41467-021-25343-7.

116. Sun, W., Dacek, S.T., Ong, S.P., Hautier, G., Jain, A., Richards, W.D., Gamst, A.C., Persson, K.A., and Ceder, G. (2016). The thermodynamic scale of inorganic crystalline metastability. Sci. Adv. *2*, e1600225. https://doi.org/10.1126/sciadv.1600225.

117. Jang, J., Gu, G.H., Noh, J., Kim, J., and Jung, Y. (2020). Structure-based synthesizability prediction of crystals using partially supervised learning. J. Am. Chem. Soc. *142*, 18836–18843. https://doi.org/10.1021/jacs.0c07384.

118. Kim, E., Huang, K., Saunders, A., McCallum, A., Ceder, G., and Olivetti, E. (2017). Materials synthesis insights from scientific literature via text extraction and machine learning. Chem. Mater. *29*, 9436–9444. https://doi.org/10.1021/acs.chemmater.7b03500.

119. Roscher, R., Bohn, B., Duarte, M.F., and Garcke, J. (2020). Explainable machine learning for scientific insights and discoveries. IEEE Access *8*, 42200–42216. https://doi.org/10.1109/ACCESS.2020.2976199.

120. Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., et al. (2020). Explainable Artificial Intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf. Fusion *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012.

121. Hoffman, R.R., Mueller, S.T., Klein, G., and Litman, J. (2018). Metrics for Explainable AI: Challenges and Prospects. arXiv preprint (arXiv). 1812.04608.

122. Wagner-Mohnsen, H., and Altermatt, P.P. (2020). A combined numerical modeling and machine learning approach for optimization of mass-produced industrial solar cells. IEEE J. Photovolt. *10*, 1441–1447. https://doi.org/10.1109/JPHOTOV.2020.3004930.

123. Pan, J., Low, K.L., Ghosh, J., Jayavelu, S., Ferdaus, M.M., Lim, S.Y., Zamburg, E., Li, Y., Tang, B., Wang, X., et al. (2021). Transfer learning-based artificial intelligence-integrated physical modeling to enable failure analysis for 3 nanometer and smaller silicon-based CMOS transistors. ACS Appl. Nano Mater. *4*, 6903–6915. https://doi.org/10.1021/acsanm.1c00960.

124. Liu, S., Kappes, B.B., Amin-ahmadi, B., Benafan, O., Zhang, X., and Stebner, A.P. (2021). Physics-informed machine learning for composition – process – property design: shape memory alloy demonstration. Appl. Mater. Today *22*, 100898. https://doi.org/10.1016/j.apmt.2020.100898.

125. Kuhn, H.W., and Tucker, A.W. (1953). Contributions to the Theory of Games*2* (Princeton University Press), p. 307.

126. Lundberg, S.M., and Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions*30* (Advances in neural information processing systems), pp. 4768–4777.

127. Ribeiro, M.T., Singh, S., and Guestrin, C. (2016). Why Should I Trust You? (ACM).

128. Kumar, I.E., Venkatasubramanian, S., Scheidegger, C., and Friedler, S. (2020). Problems with Shapley-Value-Based Explanations as Feature Importance Measures. International Conference on Machine Learning (PMLR), pp. 5491–5500.

129. Phillips, R.L., Chang, K.H., and Friedler, S.A. (2018). Interpretable Active Learning.

130. Slack, D., Friedler, S.A., Scheidegger, C., and Roy, C.D. (2019). Assessing the Local Interpretability of Machine Learning Models (NeurIPS Workshop on Human-Centric Machine Learning).

131. Kong, C.S., Luo, W., Arapan, S., Villars, P., Iwata, S., Ahuja, R., and Rajan, K. (2012). Information-theoretic approach for the discovery of design rules for crystal chemistry. J. Chem. Inf. Model. *52*, 1812–1820. https://doi.org/10.1021/ci200628z.

132. Naik, R.R., Tiihonen, A., Thapa, J., Batali, C., Liu, Z., Sun, S., and Buonassisi, T. (2021). Discovering Equations that Govern Experimental Materials Stability under Environmental Stress Using Scientific Machine Learning*8* (NPJ Comput Mater), p. 72. https://doi.org/10.1038/s41524-022-00751-5.

133. Tercan, H., Guajardo, A., and Meisen, T. (2019). Industrial Transfer Learning: Boosting Machine Learning in Production. IEEE 17th International Conference on Industrial Informatics (INDIN) (IEEE), pp. 274–279. https://doi.org/10.1109/INDIN41052.2019.8972099.

134. Ruder, S., Vulić, I., and Søgaard, A. (2019). A survey of cross-lingual word embedding models. J. Artif. Intelligence Res. *65*, 569–631.

135. Day, O., and Khoshgoftaar, T.M. (2017). A survey on heterogeneous transfer learning. J. Big Data *4*, 29. https://doi.org/10.1186/s40537-017-0089-0.

136. Mehr, S.H.M., Craven, M., Leonov, A.I., Keenan, G., and Cronin, L. (2020). A universal system for digitization and automatic execution of the chemical synthesis literature. Science *370*, 101–108. https://doi.org/10.1126/science.abc2986.

137. Tshitoyan, V., Dagdelen, J., Weston, L., Dunn, A., Rong, Z., Kononova, O., Persson, K.A., Ceder, G., and Jain, A. (2019). Unsupervised word embeddings capture latent knowledge from materials science literature. Nature *571*, 95–98. https://doi.org/10.1038/s41586-019-1335-8.

138. Olivetti, E.A., Cole, J.M., Kim, E., Kononova, O., Ceder, G., Han, T.Y.-J., and Hiszpanski, A.M. (2020). Data-driven materials research enabled by natural language processing and information extraction. Appl. Phys. Rev. *7*, 041317. https://doi.org/10.1063/5.0021106.

139. Weston, L., Tshitoyan, V., Dagdelen, J., Kononova, O., Trewartha, A., Persson, K.A., Ceder, G., and Jain, A. (2019). Named entity recognition and normalization applied to large-scale information extraction from the materials science literature. J. Chem. Inf. Model. *59*, 3692–3702. https://doi.org/10.1021/acs.jcim.9b00470.

140. Vaucher, A.C., Zipoli, F., Geluykens, J., Nair, V.H., Schwaller, P., and Laino, T. (2020). Automated extraction of chemical synthesis actions from experimental procedures. Nat. Commun. *11*, 3601. https://doi.org/10.1038/s41467-020-17266-6.

141. Swain, M.C., and Cole, J.M. (2016). ChemDataExtractor: a toolkit for automated extraction of chemical information from the scientific literature. J. Chem. Inf. Model. *56*, 1894–1904. https://doi.org/10.1021/acs.jcim.6b00207.

142. Hawizy, L., Jessop, D.M., Adams, N., and Murray-Rust, P. (2011). ChemicalTagger: a tool for semantic text-mining in chemistry. J. Cheminformatics *3*, 17. https://doi.org/10.1186/1758-2946-3-17.

143. Schwenker, E., Jiang, W., Spreadbury, T., Ferrier, N., Cossairt, O., and Chan, M.K.Y. (2022). EXSCLAIM!–An Automated Pipeline for the Construction of Labeled Materials Imaging Datasets from Literature (Computer software). https://github.com/MaterialEyes/exsclaim.