Overcoming Roadblocks for Implementing AI/ML Methods for Materials Advancement

James A. Warren (0000-0001-6887-1206),[1] Francesca Tavazza (0000-0002-5602-180X) ,[2] Austin McDannald (0000-0002-3767-926X),[2] A. Gilad Kusne (0000-0001-8904-2087),[2] Howie Joress(0000-0002-6552-2972),[2] David P. Hoogerheide (0000-0003-2918-1469),[3] Brian L. DeCost (0000-0002-3459-5888),[2] Kamal Choudhary (0000-0001-9737-8074),[4,5,6] Debra J. Audus (0000-0002-5937-7721)[4]

[1]Material Measurement Laboratory, National Institute of Standards and Technology, Gaithersburg, MD, USA, 20899
[2]Materials Measurement Science Division, National Institute of Standards and Technology, Gaithersburg, MD, USA, 20899
[3]Center for Neutron Research, National Institute of Standards and Technology, Gaithersburg, MD, USA, 20899
[4]Materials Science and Engineering Division, National Institute of Standards and Technology, Gaithersburg, MD, USA, 20899
[5]Department of Materials Science and Engineering, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD 21218, USA
[6]Department of Electrical and Computer Engineering, Whiting School of Engineering, Johns Hopkins University, Baltimore, MD 21218, USA

Xxxx. Xxx. Xxx. Xxx. YYYY. AA:1–30

https://doi.org/10.1146/((please add article doi))

Keywords

Autonomous Experimentation, Self-Driving Laboratories, AI, Machine Learning, Materials Research and Development

Abstract

The development of novel materials with tailored properties is a complex, multi-objective optimization problem that has long been a challenge in materials research. The integration of artificial intelligence (AI) and machine learning (ML) techniques has shown great promise in accelerating materials discovery, design, and development by uncovering hidden correlations between processing, structure, and properties. Autonomous experimentation (AE) platforms, also known as self-driving laboratories (SDLs), have emerged as a powerful tool in this endeavor, enabling the rapid and efficient acquisition of critical data through a closed-loop feedback process. In this review, we explore the applications of AI/ML techniques to materials research and development through the lens of SDLs, and examine the challenges and opportunities associated with the development and deployment of SDLs. We provide a detailed analysis of the components of an SDL, including AI-driven decision-making, experimental data generation, and knowledge representation, and discuss the current barriers to industrial adoption.

Contents

1. Overview

Materials research seeks to develop the "best" material for a specific application. In general, multiple properties cannot all be optimal simultaneously, but instead the designer must endure various trade-offs around cost, performance, manufacturability, and any number of specific physical properties (*e.g.*, strength, hardness, thermal conductivity, color, bandgap, and countless others). Thus, materials research and development (R&D) can be characterized as an enormously complex optimization problem, as the researcher seeks to design a fit-for-purpose material with a host of desired properties.

The rise of AI/ML approaches as a method to address this complex, multi-objective optimization problem is perhaps unsurprising, as it allows a researcher to uncover and take advantage of unrecognized correlations between processing, structure, and properties. AI can also be used to significantly accelerate property evaluation and suggest which experiment or simulation to perform next to maximize knowledge generation. There are many areas of materials research where AI methods can be brought to bear, including rapid characterization of microstructures, elucidation of trends in systems with large numbers of process variables, the development of fast surrogate models, and, of course, accelerated synthesis of new materials. In this review, we will explore a wide range of applications of AI techniques to materials R&D, describe the progress that has been made to date, and detail the roadblocks that must be overcome to realize the full promise of these methods.

Automation(1, 2), which has many benefits in its own right (including low down time, high repeatability, high metadata capture), is a natural pairing with AI-driven materials research methods, allowing for agile collection and digitization of consistent experimental data that can be fed back into a scientific AI agent to improve predictions and other inferences(3). The term "autonomous" is used to describe such a system with hardware (or physical simulation tools) controlled by an AI agent, and is therefore capable of making decisions without direct human intervention. In the scientific domain, autonomous experimentation (AE) platforms can be used to rapidly and efficiently acquire the most critical data to accelerate materials discovery, design, and development. These platforms, colloquially referred to as self-driving laboratories (SDLs), are typically composed of several tasks in a closed feedback loop, as laid out conceptually in Fig. 1(4).

This feedback loop begins with a human scientist defining an objective, specifying constraints, and providing prior knowledge to the SDL. The main AI portion of the SDL then generates a surrogate model to predict materials properties over some range of input parameter space. The AI then imparts an acquisition function on these predictions to decide what new data will help the platform achieve its objective most efficiently. This acquisition function may be tuned for some combination of exploration or optimization, depending on how the scientist configures it. The SDLs hardware will then execute a series of actions to collect this data through automated materials processing and characterization. Theory-based simulations may also be employed to generate data. This execution is often complex and requires orchestration to coordinate the various tasks. The data generated by the hardware are then analyzed and stored, which includes reducing, analyzing, and archiving the data and metadata. One may also look for errors in the data from a variety of experimental issues. These data are then fed back into the AI model to improve its predictions, and the process begins again. After a sufficient number of iterations of this loop, the human scientist can use the data and surrogate model to glean deeper insight, as well as share this information. Ideally, newly discovered materials can be applied to improve or enable new technologies.

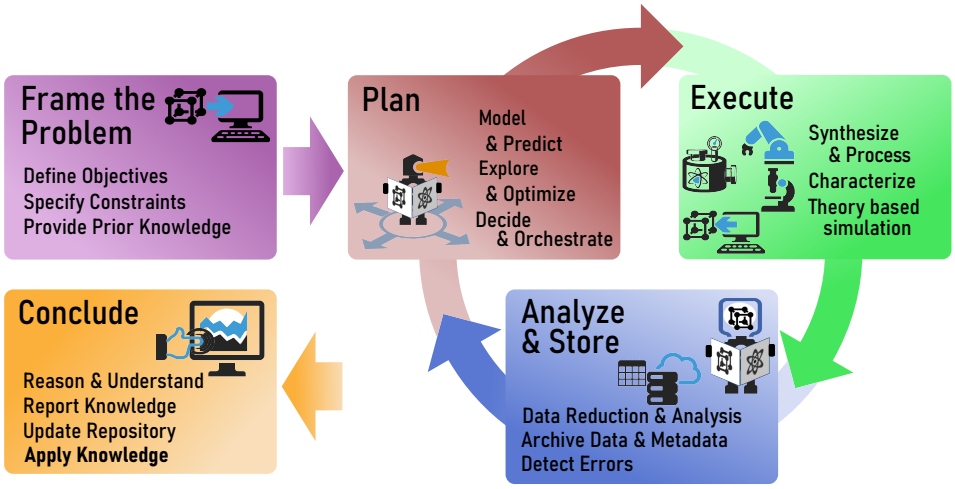

Figure 1

A conceptual view of tasks in a self-driving laboratory. Adapted from (4).

In this paradigm, the components of an SDL act as a microcosm of materials research, mimicking many of the actions a human researcher takes. These actions include both scientific/engineering thinking and experimental data generation. Both an SDL and a human need to be able to use physical principles, existing data, and new data to create a hypothesis that can be tested (and improved on) by additional experimental data, and then identify the most impactful experiment(s) to carry out. Similarly, an SDL and a human researcher must be able to process and characterize materials (or perhaps simulate them) to generate new data which can be used, in context, to improve understanding of materials and their properties. SDLs are a particularly well-suited testbed for various AI-based scientific algorithms, as the lack of human intervention means that each constituent algorithm must be highly trustworthy and accurate. Overall, the exercise of detailing the elements of an SDL research system is exceptionally useful, as it forces one to break down the research process systematically.

While much progress has been made towards the development of a wide range of AI algorithms for a range of materials science problems, there still exist a variety of challenges in achieving the full promise of scientific ML. In this article, we will explore these challenges and paths to overcome them through the lens of a SDL paradigm. While these challenges are particularly relevant in the context of an SDL, they can also frustrate research performed using standard approaches. In addition to these AI challenges, we will also discuss some of the challenges and paths forward for improving SDLs. We begin our discussion with a quick overview of the components of an SDL, and then in subsequent sections we do a deep dive into all of the components. We conclude with a discussion of the current barriers to industrial adoption, followed by a summary of the main points detailed herein.

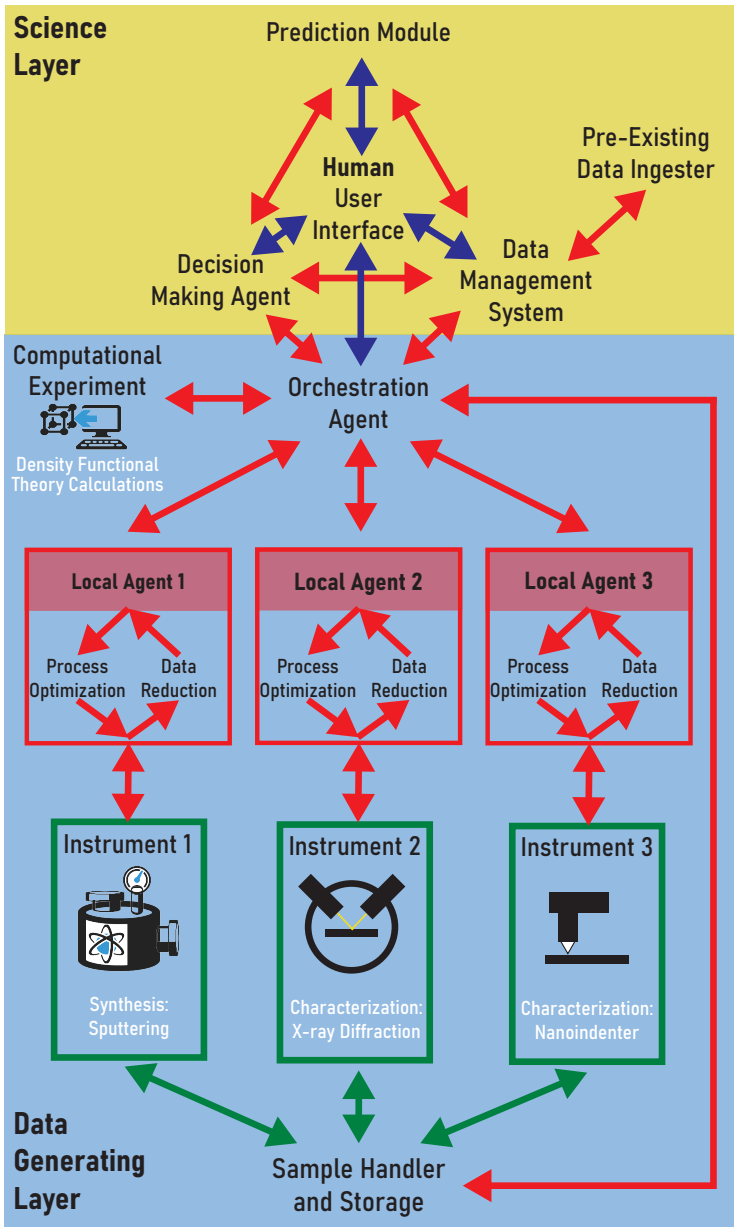

Figure 2

A potential modularization of an autonomous ecosystem with each node being a component and each arrow representing a necessary interchange standard (Red: programmatic interface; Green: physical interface; Blue: user interface). Each component can be arbitrarily complex with additional features as long as it carries out the task required of it in some way, and it may have additional connections as needed (though necessarily not fewer connections). This figure shows 3 instruments and one computation experiment, though more or fewer may be included in a particular instantiation.

2. Overview of Self-Driving Laboratories

As discussed above, an SDL must carry out all of the functions of a traditional human researcher. Fig. 2 illustrates one potential way these tasks can be discretized into modules that can be composed to create an SDL, following Joress et al. (5). While these are shown as discrete modules here, in some (perhaps even many) cases, these tasks may be combined or expanded to make use of more advanced approaches. We present a high-level overview of each of these modules within the larger context of SDLs and then focus on each of them, in turn, in the sections 3 and 4.

The SDL can be divided into two parts: the *science layer* and the *data generating layer*. The former is responsible for the scientific/engineering thinking of the platform: applying knowledge and data to generate a hypothesis (often in the form of property predictions) then creating a plan to test and improve that hypothesis. The *data generating layer* is responsible for carrying out the experimental plan (through both physical and computational experiments), to create the informative data needed by the *science layer* to improve its hypothesis.

2.1. Science layer

The *science layer* consists of 4 main parts. First, a human scientist interacts with the system, through a human user interface, to provide constraints and objectives for an experimental campaign, along with prior knowledge, typically in the form of pre-existing databases, physicochemical laws, or other known relations. These data, along with any relevant data generated by the platform (and often snapshots of model predictions) are stored in the *data management system*. Next, the *prediction module* takes the data and knowledge and generates predictions of materials properties across the composition and processing domain of interest. This is typically done with an AI-based surrogate model that can be retrained and make predictions roughly on the timescale of data generation or faster. Those predictions are then given to the *decision making agent*, which uses the predictions and their uncertainties to select a set of priority experiments to fill in knowledge gaps. This may include optimization of one or more properties, improved knowledge of a property landscape, or testing of some scientific hypothesis.

2.2. Data Generating Layer

The *data generating layer* consists of a number of instruments, along with an *orchestration agent* and a *sample handler and storage*. A *local agent* acts as an interface between the physical instrument and the rest of the SDL

The *orchestration agent*'s task is to take the list of priority experiments and coordinate the various instruments to generate those data. This includes managing instrument scheduling to maximize uptime and throughput, managing sample and data exchange between instruments, and monitoring instrument resource levels, error resolution, and other maintenance needs. The instruments then carry out their specific tasks as delegated by the *orchestration agent*.

An autonomous platform may have one or more instruments to carry out the necessary experimental actions. The instruments are typically designed to synthesize (*e.g.*, chemical synthesis platform, vapor deposition tool for thin films, additive manufacturing platform for alloys) or otherwise process materials (*e.g.*, mixing, various types of furnaces, device

fabrication, or mechanical processing) or to characterize them (*e.g.*, microscopy, diffraction, mechanical testing, electrical testing). Many instruments might intrinsically carry out multiple tasks (*e.g.*, a chemical synthesis platform might weigh out the final product to determine reaction yield, a vapor deposition system might have an electron diffraction tool as part of it, a microscope may have in situ annealing capabilities). Instruments often have a *local agent* between them and the *orchestration agent*. The fundamental purpose of this agent is to act as an interpreter between these two components – communicating the experimental directions to the instrument in the language it is expecting and outputting the data to the orchestration layer in the form it is expecting. For the former, this may involve some amount of process optimization (*e.g.*, selecting deposition parameters that are best for a given target material, tuning a microscope to optimize between resolution and field of view). Often in materials science, data are not simply a scalar but some higher dimensional representation. This representation may be difficult for the *prediction module* to directly use, often it is best to reduce these data to some set of representative scalars, through data analysis and reduction.

The instruments themselves are similar to many found in conventional materials science laboratories with the caveat that, to the extent the platform is to be operated in an automated fashion, they must operate without human intervention. This includes the ability to load and unload samples (in holders) and to receive operational commands programmatically. One important consideration is the required form factor of a material for a particular operation. For instance, while x-ray diffraction (XRD) on a bulk material requires limited sample preparation, a tensile test on that material would typically require a "dogbone" shaped specimen that needs to be prepared. If the sample synthesis does not natively generate such a form factor then specialized hardware may be required modify the sample (*e.g.*, Zhang et al. (6)).

In addition to the individual instruments, there must be hardware to handle transportation of samples between instruments and storage of samples that are not actively being used. This hardware may include motion stages, robotic arms (mounted on mobile platforms in some cases(7)), and conveyor belts.

In addition to generating data through physical experiments, data can also be generated though computational efforts. The *prediction module* carries out materials computations, typically done using models that can retrain and predict rapidly. Some computational approaches, like Density Functional Theory (DFT) and classical molecular dynamics, can produce highly useful materials predictions but are too slow to be included natively in the *prediction module*. Instead, they are be considered *computational experiments*, acting like instruments to generate data, just without the need to generate physical samples.

2.3. A review of existing SDLs

In the last decade there has been a large and growing effort to devise SDLs for a variety of materials applications. Early examples, such as the NIST scanning droplet cell (9) and the Air Force Research Lab's ARES system for testing carbon nanotube catalysts (10) were systems cleverly designed to couple synthesis and characterization with minimal sample manipulation. A great number of chemistry and liquid-phase centered SDLs were also developed, many using platforms and other standards developed for the pharmaceutical and chemical industry(8, 11, 12). Platforms that make thin films and devices from these liquid chemistries have also been developed (13, 14). In the inorganic space, the complexities of

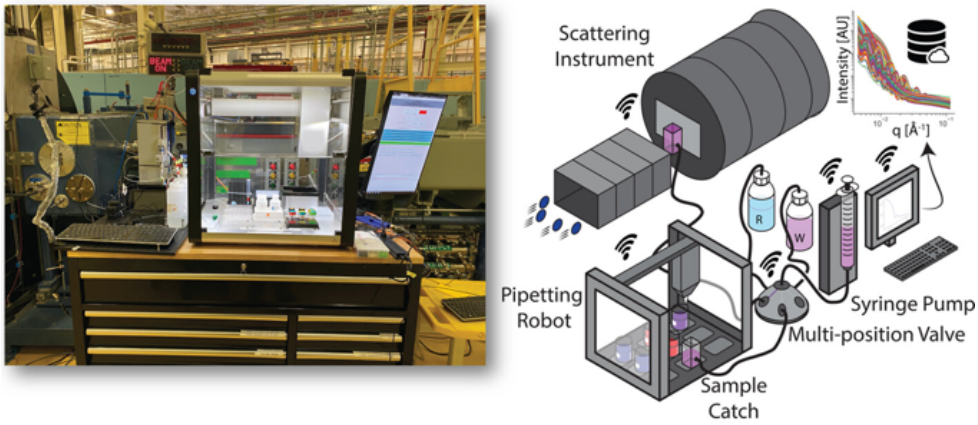

Figure 3

Photograph and schematic for the advanced formulations laboratory (AFL)(8). A pipetting gantry robot is used to prepare mixtures that are then pumped, via syringe pump, into a sample cell on a beamline (x-ray or neutron).

materials processing has made the development of SDLs much slower. Despite the expense of thin-film processing equipment, sputtered thin-films have proven to be a tractable approach (15, 16). SDLs build around powder processing have also been demonstrated including the A-lab (17) and a mobile robot based platform (18). For bulk metallurgy, the HT-READ platform has demonstrated a high level of automation for sample synthesis through additive manufacturing along with a suite of automated processing and characterization tools (19).

3. Science Layer

3.1. Data Management System

Having detailed the elements of an SDL system, we now dive into an in-depth examination of those elements. As shown in Figure 2, the *data management system* is connected to many other elements in the *science layer*. Of these connections, the most important are providing the data that the *prediction module* needs to make predictions, storing the data from the *data generating layer*, and providing data to the *human user interface*. Additionally, the *data management system* may need to handle heterogeneous data with different forms such as single values, vectors, matrices, images and from different sources including various instruments and*computational experiments*. Several different types of data storage frameworks have been used ranging from common python packages such as Pandas (20, 21) and Xarray (22) to MongoDB and Structured Query Language (SQL). Although there are an increasing number of software developed packages for SDL data management, (23, 24, 25), spinning up a *data management system* is still a challenge. In the future, common practices may eventually be adopted.

Choosing a particular data storage framework is not sufficient. Careful thought is also needed to determine what metadata needs to be stored, how the metadata is represented, how to handle both raw and processed data, and, in many cases, how and which data will be made public. In general, chemistries should be represented by common formats

such as Simplified Molecular Input Line Entry System (SMILES) (26) or, in the case of polymers, BigSMILES (27). To guide this process, it is helpful to consider all aspects that are required in order to make data Findable, Accessible, Interoperable and Reusable (FAIR). In particular, Box 2 of Wilkinson *et al.* is a useful guide. (28). Developing schemas, ontologies and data models is still difficult because community consensus and standards are required along with flexible frameworks (29) that can handle data from new measurement techniques and a diversity of uses.

3.2. Pre-existing Data

Data generated by the platform can be supplemented with relevant pre-existing data. Such datasets may be of properties that are related to properties of interest, known as "proxy" properties. Another possibility (17) is to leverage large computational datasets, such as the Materials Project (30) A third possibility is that a dataset exists for the property of interest, but with measurements carried out using different settings or different processing histories or on different but related materials. For example, this approach was used by Wu *et al.* (31) to accelerate the autonomous synthesis of electrochromic polymers with new color values via the PolyBot platform.

Using pre-existing data does create challenges as well. The first challenge is locating useful datasets. To assist readers, we direct them to a curated list of useful datasets (32). In many cases, data extraction is required. Historically, this has been mainly a manual pursuit. However, automated materials data extraction continues to improve and is likely to continue to do so as large-language models (LLMs) continue to advance. (33, 34) Even if LLMs are not able to complete every step in the data extraction process, such as finding specific values, AI including LLMs, and in particular Retrieval-Augmented Generation (RAG) LLMs that make use of external data or resources, can help identify promising sources and relevant sections for manual curation. Even in the best case, extracting datasets still requires the hard work of determining the appropriate schema which covers all the necessary metadata.

It should be noted that an inherent problem with LLMs is "hallucinations", which is a subject of study unto itself (35). Such hallucinations can be the result of both the statistical model of word prediction and erroneous training data. Indeed, all AI/ML models have inherent uncertainty. The uncertainty can be reduced by having the LLM report references that can be verified, rerunning the query a number of times, or using several models and voting for the most common suggestion. A surefire way to reduce the incidence of specific types of hallucination is by adding deterministic layers/filters to the model, but at the cost of reducing the creativity of the model. In some cases this will be desirable (for instance, adding checks to ensure that citations generated by an LLM are to real manuscripts), but in others the benefits of a statistical approach will outweigh the risks, such as guessing the intent, through a chatbot, of a researcher seeking the best methods to conduct an experimental campaign. Researchers who incorporate LLMs into their workflows need to be cognizant of these risks so they can decide which tasks are best suited to this technology. Often the risk/reward proposition to using an LLM will boil down to the costs of uncertainty, which can range a small amount of from wasted time, to reputational damage, to deleterious impacts on human health and safety.

After data extraction via LLM or otherwise, the data need to undergo "cleaning." There are several key questions that must be answered during the data cleaning stage. Is all the essential metadata present? Is the methodology appropriately defined? Is uncertainty

provided? Is the source (or subsource in the case of federated datasets) reliable? How are missing information or values handled? Is preprocessing required to remove outliers? Should data be summarized if multiple values are present, *e.g.*, by taking the mean or median? The answers to all of these questions matter and should always be considered in light of the signal-to-noise ratio for a given application. For additional information we refer the reader to a best practices guide for materials and AI. (36)

After the data have been appropriately wrangled, the final step is to make it usable by SDLs. This could involve Application Programming Interfaces (APIs) or setting up a local database. Alternatively, data sets can be used to train surrogate AI models, which natively handle missing data points. In some cases, additional steps are required to ensure the data are beneficial for the *prediction module*. If the data are from simulations, it may be systematically biased, *e.g.* Refs. 37, 38. In both this case, and that of proxy properties, (39) transfer learning (40) can be incorporated into the *prediction module*. Even if the data are for the same property and are experimental, one can assign different datasets categorical values to emphasize they are not from the same source and thus likely have some different unknown metadata (41).

3.3. Prediction Module

At the top of our example SDL workflow is the *prediction module*. This is the AI module responsible for answering the scientific question at hand. What questions are of interest informs which AI models will be useful. And, the scope of these questions has been steadily increasing. AI for materials science applications is still in its infancy and existing techniques are able to answer questions like: "Given this ternary composition what is the optimum of some property?"(42, 43, 44) or "What is the magnetic ordering behavior of this material?" (45). Baked into these implementations are caveats of particular measurement techniques, or particular material classes to consider. One step in sophistication of these techniques using multiple sources of information (sometimes at different locations of the search space) to answer the question. The approach referred to by the acronym SAGE (44) takes advantage of the well-known existence of structure-property relationships to discover phase and property maps or find optimal materials. For example, if a step change in the structure should result in a discontinuity in the properties, and the reverse is also true. Many of the generative AI techniques are also at this first step of sophistication—searching across their internal representation space describing materials to optimize some property or set of properties. These generative AI models take as input some random (or pseudo-random) number as well as a query, and are biased by their training procedure to output a hopefully useful material, such as in Takahara et al. (46).

A prerequisite for all of these AI-powered workflows—especially in the sciences—is being able to trust and interpret the AI algorithms. One way to accomplish this is to design the AI models that have some native intuitive meaning, either by having designs that are inspired by or directly represent the processes being studied. Examples of this include directly modeling the atoms, bonds, and bond angles with the graph structure of a GNN (47), using physical models (48), enforcing physical constraints (49), or incorporating physics-based loss functions. The more black-box AI algorithms at the other end of the interpretability spectrum need to be inspected with various tools in order to understand their behavior. A increasingly common approach is to use Shapely Additive ExPlanations (SHAP), which uses a game-theoretic approach to identify the contribution of each feature

to the outcome (50). This approach works best for regression tasks with a small number of predictions. For computer vision tasks there are tools, like Class Activation Maps, that allow for the visualization of what features were important at which layers of a convolutional neural net.

But for other tasks, new tools are needed to understand the behavior of the models. For instance, in some cases it is not obvious to what extent a model is extrapolating (51). This is partially due to the fact that what is considered in distribution or outside the distribution of the training data partially depends on the internal representation of the data within the model, which may only happen after training. In a study predicting formation energies of materials, Li et al. (52) showed that leaving any materials from a whole space group, or leaving out any materials that contain any elements from entire rows or columns from the periodic table does not mean the model is extrapolating to make predictions on those left out materials. This is perhaps a somewhat surprising and counterintuitive result, that their model could, for example, accurately predict the formation energy of an Al-containing material without training on any Al-containing materials or even without training on any materials that contain elements from Group 13 of the periodic table. Yet, despite this, they also showed that there is something about Materials Project 2021 that extrapolates away from Materials Project 2018 (51). The field needs new tools that can show *a priori* when a new test data point is sufficiently outside the model's own internal representation distribution that it should be considered extrapolative rather than interpolative. Recent studies investigating the extrapolative performance of ML models consider not only metrics of similarity of the test point to the training set, but also the variance, and uncertainty or confidence levels of the model predictions. (53, 54)

3.4. Decision Making Agent

Once you have a model that is making predictions, the next step in the autonomous workflow is to decide what new data, if any, to acquire next. This decision is strongly influenced by the goals of the autonomous campaign. At perhaps the most basic level, the goal could be to learn the function across the whole domain or to optimize some property or set of properties within that domain. For mapping a function across the domain, one could consider which points in the domain have the most uncertain predictions and measure at those locations. Forecasting methods, like Knowledge Gradient(55, 56) and Risk Minimization(57), pose the question as: "If our predictions are correct (we observe what the model predicts we would observe), which measurement would lower the overall uncertainty?" Thus, these methods try to find the point to measure that would minimize the global uncertainty by forecasting a prediction of that measurement and estimating the effect of that forecasted measurement on the uncertainty landscape.

If the goal is to instead find an optimum (*e.g.*, the material composition, or processing condition that optimizes some property), then other decision making algorithms are useful. These will typically use the optimum as predicted from the model, but also have some mechanism for incorporating uncertainty about that prediction. Upper Confidence Bound, for example, doesn't choose the point of the predicted optimum, but rather choose the point where the, say, 95 % confidence interval is maximized. Thompson Sampling constructs a statistical process of the predictions across the domain, samples from that process, and chooses the point that where that sample is optimized. Expected improvement also constructs a statistical process across the domain, then truncates this process below the

current observed optimum, evaluates the expected values of those truncated distributions, and picks the point where that is optimized. Often there is more than one parameter to optimize. In some cases, multiple types of parameters are fed into a figure of merit, or some other heuristic, in which case the multi-objective optimization problem collapses to a single objective problem and the aforementioned methods apply. However, in many cases the multiple parameters represent truly separate objectives, in which case the goal is to find the Pareto front - the set of observations that define the optimal tradeoff between the parameters, where at every point improving one parameter sacrifices the others. Some techniques for discovering the Pareto front include randomly weighting the objectives at each iteration,(58) or simply reverting to pure exploration and calculating the Pareto front *a posteriori*.(59) Other techniques consider the currently Pareto front of the currently observed data and seek to maximize the density of points,(60) the hyper volume,(61), or the information gain,(62) of the Pareto front.

At some higher level, the goal of an autonomous campaign might be to discover some operant mechanism or behavior. If you have several models, be they pure physics models or AI models, you could try to find the points that most distinguish between the models (the points where the predictions of each model most disagree) using an informational entropy approach.(63) That approach drives the system toward the experiment that would give you the most information about which model is best. Other statistical tools like Bayes Information Criterion (BIC) and Akaike Information Criterion (AIC) lets you compare which models are better supported by the data. Alternatively, considering the parameter-wise informational entropy can be useful in comparing models with drastically different analytical forms, or driving the experiment to improve the understanding of particular parameters of interest (64).

Perhaps one step up in sophistication would be choosing what *type* of experiment to perform. If one has a well known analytical formula that one can use to interpret each type of measurement, one might be able to directly model the information gain in choosing between those measurements (64). But what if the problem is more open-ended? When characterizing an unknown powder, should you use XRD, Raman spectroscopy, or Fourier-transform infrared spectroscopy (FTIR)? XRD provides information about the spacing between planes of atoms, Raman spectroscopy provides information about phonon modes in the material, and FTIR provides information about the local chemical environment of each atom. Given the information currently available, what's most useful measurement to perform? Such open-ended questions are the subject of ongoing research.

One of the main challenges in selecting a decision making technique is in evaluating its performance. In the absence of a well defined traditional workflow for measurement acquisition, random sampling is often a fair benchmark for autonomous decision making algorithms. Such a benchmark can provide information about the relative performance of autonomous workflows on specific problems. Perhaps more important for the widespread adoption of autonomous workflows is to understand some sense of which set of algorithms are appropriate or well suited for particular materials science and engineering research interests. The community needs a set of well defined problems, a virtual gym (*à la* the OpenAI Gym for reinforcement learning (65)), for testing different algorithmic designs. In the materials characterization applications, this could include: trying to correctly identify the phases present in a increasingly difficult set of XRD analysis problems, or trying to construct the phase maps or property maps across a compositional space.

In a future where we have autonomous systems aimed at solving grand engineering chal-

lenges or discovering new physics, how will we be able to evaluate their performance? One way might be to borrow challenges from historical discoveries. Given the resistivity behavior of a some conventional superconductors, along with a few other metals, insulators, and superconductors, how quickly can a set of algorithms discover the cuprate superconductors? How well would autonomous systems be able to replicate the historical discoveries that were at least partially aided by serendipity? Could such systems recognize an unexpected result that would change the underlying assumptions of the field? Given the atomic arrangement descriptions of a set of crystalline and amorphous materials, how well can the algorithms handle quasicrystals, which break the historical assumption that long range order required translational symmetry?

3.5. Human User Interface

Widespread adoption of autonomous systems will require integrating these systems into facility-wide, human-lead processes. Human users will need to be able to interact with these systems, including monitoring system activities, extracting knowledge from the governing AI, and potentially imparting their own knowledge to the AI. This will require a human interface. An effective interface provides easily interpretable information regarding system operations and status as well as AI-based analysis, prediction, and decision making. The interface must also provide a clear means for the user to impart their knowledge and modify operations when needed. Such an interface may facilitate user trust in the system as well as greater use and adoption.

Currently, two modes of interfacing with self-driving labs have been explored in the literature: reinforcement learning (66) and human-AI teaming. In both cases, these interfaces are bespoke to their systems. This is due to differences in equipment and driving-AI from SDL to SDL. These SDL-to-SDL differences poses a challenge to developing more generally applicable user interfaces. Additionally, while graphical user interfaces facilitate users of a wider background, interaction with the majority of SDLs is performed through either scripting or programming, requiring advanced knowledge of interaction protocols of the controlled equipment and the AI. This poses a significant challenge for SDL adoption in common manufacturing pipelines.

Interfaces dependent on reinforcement learning such as that of Ref (67) seek to identify an optimal action policy, i.e., mapping relevant situations to desired SDL actions. The interface imparts knowledge of the current action policy, and the user indicates approval or disapproval. Such methods typically require large amounts of training data to achieve adequate performance – a situation not common in research settings.

An alternative mode of human-SDL interaction is that of human-AI teaming. In such situations, it is assumed that the user and the AI are able to impart similar types of information. In Adams *et al.* (68) the SDL seeks to discover the composition-phase map of a new material system with the minimum number of structure analysis experiments (e.g., x-ray diffraction). At every stage of analysis, the human can impart knowledge similar to that of the AI, e.g., phase identification, potential phase boundary, or materials of interest. Such a human-AI teaming requires interpretable, interactive visualizations of the AI's analysis, predictions, and decision making at every iteration. In Adams *et al.*, (68) the user interacts with a phase map through simply drawing lines and regions to indicate potential phase boundaries or regions.

4. Data Generating Layer

4.1. Orchestration Agent

As mentioned above, the orchestration module is the point of interaction between the decision-making module and the physical hardware and computational tools that will generate the data. This orchestration module schedules the system operation, executes the operation, and then collects generated data. The orchestration module must also be able to navigate the diverse communication protocols across physical and computational systems. In many cases it may be designed to interact with users, to manage maintenance, sample removal, and consumable replenishment. Additionally, the orchestration module of one SDL may be designed to interact with the orchestration module of another SDL. For instance, if the SDLs are co-designed, they can be set up to communicate through a common internet of things protocol (69).

Perhaps the biggest challenge in orchestration is scheduling. In some simple systems, the process flow is linear, with a single sample being generated between iterations of the active learning loop. Many AE platforms are more advanced carrying out several experiments simultaneously, sometimes in batch processes. These experiments may have several steps with complex inter-dependencies and a range of constraints. The task of orchestration agent with respect to scheduling, is then ensuring that these tasks can occur in their proper sequence, without interfering with each other, in the most efficient way possible. This may involve ordering experiments and their sub-tasks to balance the scientific value, assigned by the decision making agent, against the temporal and resource cost of the experiments. In complex cases this may require interaction with the decision making agent. Further complications including the need to dynamically add experiments to the queue as new data is ingested into the science layer and various processes having times that may be varying dynamically. The latter may results from feedback loops within tasks that determine the completion of that task.

4.2. Local Agent

In order for synthesis, processing, or measurement systems to carry out the tasks required of them by the science layer through the orchestration layer, the local agent must be able to modify system parameters and execute the desired operation. For instance, in a scanning droplet cell system for electrochemical measurements,(9, 70), settings of the potentiostat and the composition and flow rate of the electrolyte, need to be correctly set in order to generate the correct data. Additionally, as is often the case including for combinatorial libraries(71), samples may be spatially complex; the local agent must be able to translate sample coordinates to motor positions in order to locate the probe at the relevant part of the sample.

An optimal communication protocol allows the module to interact in its native scripting environment. For example, if the AI is python based, a python application programming interface (API) (72) to target systems would reduce potential communication complexities. Less optimal communication interfaces include having the orchestration module build an external script that dictates the desired operation of the target system, and then having the module execute the script from the command line. However, often many systems require interaction through a graphical user interface. The orchestration module must then interact graphically with this interface. As interfaces can move or change, this can result in the need for significant human guidance. The most frictional communication protocol requires the

orchestration agent to interact with a physical interface. Here the module must activate motors to push buttons and turn knobs on the physical interface to set system parameters and execute operation.

While relaying these parameters into the platforms hardware a fundamental purpose of the local agent, there are several tasks more complex agents may need to carry out. Some complex instruments, such as a chemical synthesis platform (13), may require internal orchestration to function efficiently. Additionally there may be internal hyperparameter optimization and data analysis tasks, specific to the instrument, that need to be carried out. These specialized functions are described in detail in the following sections.

4.3. Local Agent: Process Hyperparameter Optimization

Most materials research processes are complex, having a variety of parameters that must be decided on and set for a given operation. An example in materials processing might be adjusting the laser power during additive manufacturing of an alloy. We can think of these as "hyperparmeters," similar to hyperparameters in AI and other models, which are not themselves primary targets of the experimental campaign but may critically affect the outcome. While in some cases these hyperparameters may remain fixed, in others they are must be tuned to achieve reasonable results – from the example above if one switches from a Mg alloy to a refractory alloy, the laser power must be changed or there will be a high degree of porosity in the printed alloy. Optimizing these hyperparameters may greatly increase the quality of the results of the process and the speed at which new materials can be discovered and developed. Optimizing these hyperparameters requires some type of outcome to feedback on (in many cases a precursor to the main optimization goal). If that feedback loop is large then most likely these hyperparameters will need to be included in the broader search space. However, if feedback can be generated more locally, then what is effectively a nested autonomous loop can be created. This is often relevant for materials characterization, since the feedback mechanism is inherent in the data being generated.

Most materials characterization techniques are quite complex, involving trade-offs across multiple axis, generally including resolution, signal-to-noise ratio (SNR), collection scope, and collection time. For instance in XRD using an area detector, there is a tradeoff when changing the sample-to-detector distance between the resolution, the range of the data collected in reciprocal space, and the solid angle (and thus SNR level). SNR can separately be tuned by lengthening the measurement time, at the expense of time resolution and measurement throughput. Traditionally, expert users of characterization instruments have the experience to tune these hyperparameters, and while less expert users use "pre-defined" value that may not be optimal. Algorithms can be developed to tune these hyperparameters based on the data being generated. In a simple case these hyperparameters can be tuned to optimize the generated raw data to meet some set of heuristics, but more advanced algorithms can take into account the details of the scientific questions being asked of the data, along with some model of time for data reduction, to optimize the data collection (and thereby the generated data) to produce the lowest uncertainties on the scientifically relevant values extracted from the raw data.

In particular, when access to a measurement resource is limited, such as at x-ray and neutron scattering user facilities, improving measurement speed is critical to increase throughput of these unique tools. This has led to active development of self-driving measurement techniques at these facilities. In this case, the goal of the autonomous experiment is to

efficiently collect a set of data that contains requested information. Examples include selecting regions of an image for spectroscopic analysis (73), determining the location of an order parameter (45), and favoring high-intensity regions in reciprocal space (74).

In each of these cases, the measurement instrument is configured to collect data; the *prediction module* then interprets these data in the context of a model, often using Bayesian inference; and a *decision making agent* uses this model to suggest measurements at new instrument configurations. Gaussian processes are frequently employed as the model; these are advantageous because they can be quite general, when knowledge of the data landscape is limited, and they can incorporate problem-specific knowledge (75).

In the special case where instrumental data are understood in terms of specific mathematical models, additional performance improvements are possible. Of the model parameters, some are typically of scientific interest, while others are incidental to the specific conditions or instrument used and warrant less attention. An example is the *AutoRefl* measurement agent developed for neutron reflectometry (NR) (64), which allows for selecting measurement conditions that specifically optimize the parameters of interest. This is a specific case where incorporating physical information into a model enables performance improvements of 50 % or more, even on a limited one-dimensional search space. *AutoRefl* also features forecasting, *i.e.* looking ahead by multiple measurement steps, to support implementation of asynchronous measurement and analysis queues.

4.4. Local Agent: Data Reduction and Analysis

Materials data are often complex, high dimensional, and highly multi-modal: images, spectra, reciprocal space data, discretized curves. Many of the recent AI applications attempt to use these data in as close to its raw form as possible. This is in part because quantitative analysis and interpretation is often a slow manual process that requires significant expertise. This analysis is also subject to a high degree of human bias through the selection of the physical models, approximations, and estimation strategies used to turn raw data into measurements.

Bridging the gap between systems that fully rely on black box predictions and systems that rely on mechanistic interpretation is an important growth opportunity in applications of AI to science. For example, AI driven modeling of XRD data often focuses on the task of structural phase identification; however this data contains much more fine-grained information about the material structure, including volume fractions, approximate crystallite sizes, and other statistical microstructure features. Providing this type of quantitative characterization to broader AI models for material processing and properties may increase the sample efficiency for optimization, and opens up new opportunities for interpretable and mechanistic AI.

Realizing these kinds of systems is a significant challenge. Two kinds of computational problems must be solved simultaneously: model search and model estimation. This could build on existing work applying symbolic regression (for example the work of Ouyang et al. (76)) and genetic programming (for example Hernandez et al. (77)) for model discovery.

4.5. Instruments and Sample Management

Together, the instruments and the sample management system comprise the hardware of an SDL. The purpose of this hardware is to carry a set of experiments that generates data which can then be put into context by AI or human scientists to generate new or improved

knowledge. In order to generate these data, these instruments require clear communication of what experiments should be carried out along with the necessary resources (*e.g.*, precursors, consumables) to achieve that result. The equipment then generates data, typically in raw form, along with samples. This may include scalar data (*e.g.*, hardness values, conductivity), data vectors (*e.g.*, XRD data, Raman and FTIR spectra, mechanical loading curves), 2D data (e.g, non-integrated diffraction patterns, microscopy images) and higher order data (*e.g.*, time resolved image stacks, spatially resolved spectroscopy data).

The ultimate goal of SDL hardware is to make it nearly effortless for humans to rapidly collect this critical materials data. Achieving this requires existing SDLs to have many types of instruments tied together into a single platform. Beyond this, the ability to easily add capabilities to SDLs is critical for their longevity. Currently, the main barrier to achieving this is the large amount of bespoke engineering typically required to create a full-scale automated laboratory, particularly in the materials science domain. Most materials research infrastructure is human-centric, designed for samples to be manipulated by humans and operated through graphical user interfaces or other physical controls. To the extent these instruments have some amount of automation enabled, they are typically through bespoke, and often proprietary, interfaces – both at the hardware level for sample interchange and software level for controls. Thus, constructing a platform typically requires a large amount of engineering specific to the instruments being agglomerated. Identifying vendors with the necessary technical breadth to work with a wide range of instrument types is challenging and procuring custom platforms is costly(78). In-house engineering of SDLs can be time consuming and similarly costly. Ultimately, as research goals and topics shift, these custom systems are too "brittle" to enable adding of new capabilities.

There are three main opportunities for achieving increased flexibility in automation of instruments. Perhaps the most actionable approach presently is the development of robotic arms and other humanoid like robotics for the materials science lab(18, 2). These types of robots can be trained to work within existing materials science labs, using equipment with minimal amount of modification. The robotics necessary for this type of interaction have become much cheaper, safer, and easier to program within the last decade(2). This type of approach has been demonstrated for a variety of materials science lab tasks including chemical synthesis, measuring powders, and performing XRD(18, 7). AI for image recognition can greatly decrease the amount of programming required many robotics tasks(79, 80). Perhaps the biggest risk to using these types of robots is the risk to humans working in or around them. These approaches often require more complex automation than more automated-instrument centric approaches (*e.g.*, using a robotic arm to transfer liquids as compared to a pump).

A second ongoing effort is the development of new instruments designed with automation in mind. There are many cases where instruments designed to be operated without humans can be designed differently. While there are many commercial vendors working to create such instruments, there is also a large push, particularly in academia and the government research sectors to create open-source hardware. This type of hardware leverages the availability of technologies such as additive manufacturing, modular construction toys, single board computers, and other types of programmable logic controllers to create experimental hardware cheaply. The plans for these hardware, including CAD files, parts lists, and control software are often published online. These types of instruments are often built with automation in mind, but if not they are typically readily modifiable as they are being constructed.

The third and perhaps biggest opportunity for easing the construction of SDLs is perhaps the creation of interchange standards between various hardware (as well as software) components of the platform. Interchange standards would abstract one instrument or other module from the rest of the platform, meaning any compatible instrument could be added with minimal amount of engineering. For instruments, these interchange standards would include physical descriptions of the sample holder to allow samples to be moved from one instrument, through the sample handling system, to another instrument. Interchange standards would also require standards for controlling instruments as well as data formats. More details can be found in Ref. Joress et al. (81, 5).

4.6. Computational Experiments

Computational experiments play a central role in modern materials design by enabling the exploration and prediction of materials behavior across multiple length and time scales. These simulations ranging from quantum-level calculations to mesoscale and continuum models provide valuable insights that complement and, in some cases, guide experimental investigations.

With the growing availability of data and compute resources, artificial intelligence (AI) and machine learning (ML) are being increasingly integrated into computational workflows to enhance efficiency, accuracy, and scale. However, fully realizing their potential requires overcoming several technical and infrastructural roadblocks.

Multi-scale modeling serves as a key approach for linking atomistic simulations to macroscopic properties. When augmented with AI/ML, this modeling framework has the potential to accelerate materials discovery and optimization substantially. Yet, significant roadblocks such as concerns about accuracy, transferability, and computational cost, continue to limit the widespread adoption of AI in computational materials science (82).

Traditional methods such as Density Functional Theory (DFT) and other electronic structure methods provide fundamental insights but are computationally expensive. AI-driven surrogate models have been developed to accelerate these calculations, yet their reliability across diverse materials systems is still under scrutiny. Some of these surrogate models include 1) fingerprint models (such as MatMiner and classical force-field inspired descriptors, CFID)(83, 84), 2) graph neural networks (such as ALIGNN (47)), and 3) language based models (such as AtomGPT) (85, 86).

While the above models can be used for property predction/classifications tasks, atomistic simulations, including Molecular Dynamics (MD) and Monte Carlo methods, often rely on force fields. Machine-learned potentials, especially based on graph neural network force-fields (such as ALIGNN-FF (87) and MatterSim (88)) have improved efficiency but still face challenges related to generalization beyond training datasets.

In addition to the above forward materials design, AI-based methods have been used for inverse design tasks including XRD patterns and scanning transmission electron microscopy (STEM) image data to atomic structure (89, 90). This is specially potentially useful as there is no direct physics based models for such tasks.

At a larger scale, continuum and phase-field models aim to capture mesoscale and macroscale behaviors, benefiting from ML-enhanced numerical solvers that predict phase transitions, microstructural evolution, and defect dynamics with improved efficiency.

One of the major challenges in integrating above computational techniques with experimental approaches is the lack of standardized datasets. Similar to the needs for benchmark-

ing described below, computational and experimental data often differ in format, quality and uncertainty (91), limiting the ability of ML models to generalize effectively. Moreover, parameterization remains an issue, as ML-driven models tend to struggle with extrapolation when faced with novel materials compositions and structures.

Open-source benchmarking platforms such as JARVIS-Leaderboard (92) and MatBench (93) provide a strong foundation for standardizing AI/ML tasks and fostering community collaboration to solve challenging problems. However, widespread community adoption of such infrastructures remains challenging, as they often require additional effort.

5. Benchmarking, Validation and Uncertainty Quantification

As with any new scientific methodology, a critical question for novel applications of AI is "How do we know that they work?" Although there is a great deal of discussion in the current literature on measuring and benchmarking the robustness and generalization performance of AI in general Alampara et al. (94), there is relatively little specific discussion on how to effectively benchmark many of the tasks surrounding the automation aspects of SDLs. Evaluating SDLs' performance can be subjective because it is a multi-objective problem. Thus, there is a strong need for the community to identify SDL's desirable future capabilities and design quantitative benchmarks to provide guidance towards measurable research progress. As always in AI-related research, it is important to avoid the pitfall of overfitting to test set / benchmark tasks.

In addition to modeling and data quality benchmarks, there is a need for granular benchmarks for components of SDLs, and systems integrations for full SDL systems. There is currently a strong emphasis on benchmarking active learning convergence rates and predictive accuracy of the underlying models, but relatively little quantitative effort has been focused on reproducibility, robustness, and the quality and calibration of the uncertainty estimates of the models underlying AI planners.

The reproducibility and robustness of SDLs encompass both software and hardware (92). While there is a growing focus on quantifying the reproducibility of modeling pipelines, assessing the reproducibility of physical synthesis and measurement systems is a significant growth opportunity. As standardized modular hardware interfaces and protocols are developed, there is a need for systematically assessing the conformity of individual components to the standardized interfaces and functionalities. This could take the form of a suite of functional tests on reference samples to ensure that automated experimental systems yield expected results. Similarly, accurate tracking of instrumental drift from initial calibration can strongly impact the ability of SDL platforms to effectively implement the control inputs requested by automation programs. Development of automated routines to identify and correct calibration drift would benefit from standardized testing protocols designed to ensure performance within tangible calibration metrics which need to be developed.

In addition to tests for individual components, rapid development and improvement of SDL platforms will benefit from systems integration style benchmarks that assess the ability of multiple subsystems to work together to consistently implement the requested experiments. This could be achieved through a suite of baseline problems, such as checking that an SDL platform is capable of experimentally reproducing a simple known result.

Finally, since uncertainty is central to most of the acquisition policies used to select experiments in an SDL, there is a need for stronger benchmarks on model uncertainty calibration, assessment of experimental uncertainties, and propagating uncertainty estimates

through multiple layers of measurement systems, ML models, and planning systems.

6. Industrial relevance - Barriers to Adoption

SDLs are being developed in a number of universities and governmental research labs with very encouraging results(95). However, transitioning from research institutions to industrial R&D is not necessarily straightforward. While SDLs promise to significantly accelerate the discovery and optimization of new materials, they are still not sufficiently reliable, reproducible, or cost-effective for investment by the bulk of industrial stakeholders. Indeed, some highly capitalized companies can (and have) made investments in this space, and there are a small number of start-ups dipping their toes into the water, but the risk-reward proposition for most companies remains daunting. In this section we will cover some of the reservations that industry currently has and pathways to address them.

In a meeting dominated by industrial R&D and instrument vendors, the XRD community (78) declared interest in implementing increased automation for sample preparation and measurement, currently a task performed largely by people. This was motivated by the desire to reduce the level of human involvement in sample manipulation, as it is repetitive, time-consuming and can involve significant safety hazards. The participants also indicated that capabilities to automatically perform error detection would be of substantial benefit. Similar interests and concerns were also the core findings in a different workshop, the 'Accelerated Materials Experimentation Powered by the Autonomous Materials Innovation Infrastructure (AMII)' workshop.(96)

Several challenges that are limiting or delaying the addition of such automated or, possibly autonomous capabilities, were also described. Among the most important are the lack of standardization for automation interfaces, lack of connectivity protocols among instruments, lack of data format standards for output data, and inability to autonomously evaluate the quality of the data produced (especially in terms of the sample preparation). Funding issues were mentioned as well, where academia and national labs have the funds to build customized laboratory equipment (partly aided by less stringent time requirements and a lower cost workforce) while small companies typically do not. Unfortunately, currently much of the knowhow generated from academic SDLs does not directly transfer to industrial cases, mostly because of a difference in goals. Academic SDLs are focused on scientific discovery - new materials, new processing-structure relationships, etc. Contrastingly, the industrial production might be more interested in maintaining process control and surfacing early failure warnings. For example, the aforementioned interest of XRD vendors to detect poor sample preparation might not be directly transferable to the SDLs designed for scientific discovery.

Another factor affecting the adoption of SDLs by industry is the decentralized development of academic software, which is often poorly written, documented, and maintained. Moreover, it relies heavily on open-source tools, and industry (including instrument vendors) is reticent to incorporate open-source products into their own software because they would need to assume the responsibility of maintaining and supporting it. Assembling specific working groups, or consortia, could substantially increase the effectiveness of academic software for industrial purposes, as they could address long term maintenance issues as well as build trust in SDL software tools through the development of tests to score the performance of new algorithms.(81)

Users of the equipment pointed out that programmatic access to the instruments, both

for control-purposes and to access measurements data and diagnostic information, is a needed step toward building SDLs. Vendors, however, were not ready to implement such changes in their equipment mostly because: 1) demand is not currently sufficiently high to justify such a possibly large investment. 2) the custom aspect of such modifications, as very different interfaces may be involved in each case. 3) uneasiness about allowing access to their machines to software not developed in-house, as it may interact negatively with other parts of the equipment or have maintenance issues in the future. Building a user facility model for autonomous systems was suggested as a possible solution, as it would provide the opportunity to develop new standards for laboratory equipment, as well as share the cost of getting started in the autonomous arena.

Another discussed key issue was the need to develop models for proprietary data while protecting those data from unauthorized disclosure. In general, intellectual property issues are extremely important to industrial customers. They may be a major issue when setting up collaborations with academic institutions because academia's mission is to generate public knowledge (publishing in scholarly journals, for instance) while industry is interested in developing an advantage over their competitors. Creating consortia including academia and industry was the most suggested approach to overcome this challenge.

A last issue brought up in these reports was the lack of qualified personnel to run such SDLs. As only a limited number of professionals currently have expertise in materials science, AI/ML, and automation technologies, there is a need for more across-discipline training. Grants to universities to start such multidisciplinary courses were the most supported approach to address this important issue.

7. Summary

Herein we have explored the challenges associated with the application of AI to materials R&D through the lens of Self Driving Laboratories. While we used SDLs as an example, the paradigm for how an SDL functions is really just a recapitulation of how R&D is performed, whether by humans or machines, with AI algorithms and robots substituting for researchers where feasible. Thus many, if not most, of the challenges identified are not unique to SDLs.

More generally, the successful application of AI to materials R&D is, unsurprisingly, dependent on the quality of available data and the lack of uncertainty quantification in much of the published data. Some of this can be mitigated by building up a data infrastructure. Yet, this remains challenging even though a number of platforms have arisen to address this, but their use is far from common. At the bench-level each researcher must determine what data should be stored, and what discarded, and how it should be represented, both for internal use and for publication. Ultimately, representation is a standards question, but whether these standards arise from technical discussions or more organically will be discipline dependent for the foreseeable future. We would suggest that the most promising approach is through the creation of user-friendly software platforms where AI-research can be performed, and data curation is a necessary condition for deposition onto these platforms, which can then be used by subsequent researchers after some embargo period.

There are several deep mathematical issues that will continue to present challenges for the materials researcher. First, the lion's share of existing AI approaches lack interpretability and are opaque to the causal links between relevant variables. Thus, there is a pressing need for interpretable AI algorithms. Second, there is an inherent lack of invertiblity, that is, the so-called "inverse problem" where the properties and performance are specified and the

processing-structure are inferred. This has been a long-standing challenge, and systematic approaches to addressing it will continue to be in demand.

We also explored the challenges in AI-Human teaming, which are deeply linked to interpretability as the human requires interpretable visualizations of the AI's analysis, predictions, and decision making at every iteration. The user must understand the AI's current beliefs before being able to impart their own knowledge to the system.

The adoption of AI by the community will progress in the usual way new technologies infiltrate any system, first as piecemeal insertions of AI algorithms, but ultimately a more comprehensive approach will transform the materials R&D landscape. What we have right now are a bunch of AI-based solutions to various pain points but having a well-developed ecosystem of AI-based tools allows for a reimagining of how the research is done, as is exemplified in autonomous systems.

The lack of well-characterized challenge problems for materials science applications is another obstacle hindering not only AI-model development and decision-making, but hindering the deployment of autonomous systems. If the community had a "virtual gym," AI-models and decision-making algorithms could be developed against tests with practical meaning, which also would serve to help demonstrate the efficacy of autonomous systems to any organizations looking to invest in autonomous research systems. Similarly, the availability of benchmarking data allows algorithm developers to test their models, and develop confidence in their methods. Simultaneously, it allows users of algorithms to identify the best algorithms, which is more efficient.

Somewhat particular to SDLs, we explored industrial adoption, which, while these approaches are enormously promising, adoption is impeded due to a number of issues, including lack of interoperability standards, and the usual hesitancy to be a "first mover" in a dynamically evolving field. To de-risk these issues there was enthusiasm for some form of user facility, which would allow practitioners to debug interoperability problems and evolve standards. Additionally, for industry to succeed in employing these approaches, a well-trained workforce will be required, and is tacit in the above discussion.

While this article has delved into the many challenges associated with applying AI to materials R&D, we are certain that the transformation of research that these approaches allow, will have profound implications for the entire enterprise of materials research. We are now entering an age of massively accelerated materials discovery, design and deployment through the application of AI, especially through the means of Self-Driving Laboratories.

8. Disclaimers

Certain equipment, instruments, software, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement of any product or service by NIST, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

These opinions, recommendations, findings, and conclusions do not necessarily reflect the views or policies of NIST or the United States Government.

LITERATURE CITED

1. Melodie Christensen, Lars P. E. Yunker, Parisa Shiri, Tara Zepel, Paloma L. Prieto, Shad Grunert, Finn Bork, and Jason E. Hein. Automation isn't automatic. *Chem. Sci.*, 12:15473–15490, 2021. . URL http://dx.doi.org/10.1039/D1SC04588A.
2. Benjamin P MacLeod, Fraser GL Parlane, Amanda K Brown, Jason E Hein, and Curtis P Berlinguette. Flexible automation accelerates materials discovery. *Nature Materials*, 21(7): 722–726, 2022.
3. Howie Joress. Automating the materials science laboratory, 06 2025.
4. Eric Stach, Brian DeCost, A Gilad Kusne, Jason Hattrick-Simpers, Keith A Brown, Kristofer G Reyes, Joshua Schrier, Simon Billinge, Tonio Buonassisi, Ian Foster, et al. Autonomous experimentation systems for materials development: A community perspective. *Matter*, 4(9): 2702–2726, 2021.
5. Howie Joress, Brian DeCost, Katelyn Jones, A Kusne, Austin Mcdannald, Zachary Trautt, and Francesca Tavazza. Towards a composable, modular laboratory ecosystem for autonomous materials research and development, 08 2025.
6. Bojing Zhang, Leon Merker, Alexey Sanin, and Helge S Stein. Robotic cell assembly to accelerate battery research. *Digital Discovery*, 1(6):755–762, 2022.
7. Benjamin Burger, Phillip M Maffettone, Vladimir V Gusev, Catherine M Aitchison, Yang Bai, Xiaoyan Wang, Xiaobo Li, Ben M Alston, Buyi Li, Rob Clowes, et al. A mobile robotic chemist. *Nature*, 583(7815):237–241, 2020.
8. Peter A. Beaucage and Tyler B. Martin. The autonomous formulation laboratory: An open liquid handling platform for formulation discovery using x-ray and neutron scattering. *Chemistry of Materials*, 35(3):846–852, 2023. . URL https://doi.org/10.1021/acs.chemmater.2c03118.
9. Brian DeCost, Howie Joress, Suchismita Sarker, Apurva Mehta, and Jason Hattrick-Simpers. Towards automated design of corrosion resistant alloy coatings with an autonomous scanning droplet cell. *JOM*, 74(8):2941–2950, 2022.
10. Pavel Nikolaev, Daylond Hooper, Frederick Webber, Rahul Rao, Kevin Decker, Michael Krein, Jason Poleski, Rick Barto, and Benji Maruyama. Autonomy in materials research: a case study in carbon nanotube growth. *npj Computational Materials*, 2(1):1–6, 2016.
11. Tony C Wu, Andrés Aguilar-Granda, Kazuhiro Hotta, Sahar Alasvand Yazdani, Robert Pollice, Jenya Vestfrid, Han Hao, Cyrille Lavigne, Martin Seifrid, Nicholas Angello, et al. A materials acceleration platform for organic laser discovery. *Advanced Materials*, 35(6):2207070, 2023.
12. Mareike Schreiber, Manuel Brunert, and Gerhard Schembecker. Extraction on a robotic platform–autonomous solvent selection under economic evaluation criteria. *Chemical Engineering & Technology*, 44(9):1578–1584, 2021.
13. Benjamin P MacLeod, Fraser GL Parlane, Thomas D Morrissey, Florian Häse, Loïc M Roch, Kevan E Dettelbach, Raphaell Moreira, Lars PE Yunker, Michael B Rooney, Joseph R Deeth, et al. Self-driving laboratory for accelerated discovery of thin-film materials. *Science Advances*, 6(20):eaaz8867, 2020.
14. Aikaterini Vriza, Henry Chan, and Jie Xu. Self-driving laboratory for polymer electronics. *Chemistry of Materials*, 35(8):3046–3056, 2023.
15. Ryota Shimizu, Shigeru Kobayashi, Yuki Watanabe, Yasunobu Ando, and Taro Hitosugi. Autonomous materials synthesis by machine learning and robotics. *APL Materials*, 8(11), 2020.
16. Davi M Fébba, Kevin R Talley, Kendal Johnson, Stephen Schaefer, Sage R Bauers, John S Mangum, Rebecca W Smaha, and Andriy Zakutayev. Autonomous sputter synthesis of thin film nitrides with composition controlled by bayesian optimization of optical plasma emission. *APL Materials*, 11(7), 2023.
17. Nathan J Szymanski, Bernardus Rendy, Yuxing Fei, Rishi E Kumar, Tanjin He, David Milsted, Matthew J McDermott, Max Gallant, Ekin Dogus Cubuk, Amil Merchant, et al. An autonomous laboratory for the accelerated synthesis of novel materials. *Nature*, 624(7990): 86–91, 2023.

18. Amy M Lunt, Hatem Fakhruldeen, Gabriella Pizzuto, Louis Longley, Alexander White, Nicola Rankin, Rob Clowes, Ben Alston, Lucia Gigli, Graeme M Day, et al. Modular, multi-robot integration of laboratories: an autonomous workflow for solid-state chemistry. *Chemical Science*, 15(7):2456–2463, 2024.
19. Kenneth S Vecchio, Olivia F Dippo, Kevin R Kaufmann, and Xiao Liu. High-throughput rapid experimental alloy development (ht-read). *Acta Materialia*, 221:117352, 2021.
20. The pandas development team. pandas-dev/pandas: Pandas, February 2020. URL `https://doi.org/10.5281/zenodo.3509134`.
21. Wes McKinney. Data Structures for Statistical Computing in Python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 56 – 61, 2010. .
22. S. Hoyer and J. Hamman. xarray: N-D labeled arrays and datasets in Python. *Journal of Open Research Software*, 5(1), 2017. . URL `https://doi.org/10.5334/jors.148`.
23. Theophile Gaudin, Ian Benlolo, Zheng Yu Cui, Riley Hickmann, Isaac Tamblyn, and Alan Aspuru-Guzik. Molar, July 2022. URL `https://doi.org/10.5281/zenodo.6809291`.
24. Hatem Fakhruldeen, Gabriella Pizzuto, Jakub Glowacki, and Andrew Ian Cooper. Archemist: Autonomous robotic chemistry system architecture. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 6013–6019, 2022. .
25. Hatem Fakhruldeen, Gabriella Pizzuto, Jakub Glowacki, and Andrew Ian Cooper. Archemist, July 2022. URL `https://github.com/cooper-group-uol-robotics/archemist`.
26. David Weininger. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28 (1):31–36, 1988.
27. Tzyy-Shyang Lin, Connor W Coley, Hidenobu Mochigase, Haley K Beech, Wencong Wang, Zi Wang, Eliot Woods, Stephen L Craig, Jeremiah A Johnson, Julia A Kalow, et al. Bigsmiles: a structurally-based line notation for describing macromolecules. *ACS central science*, 5(9): 1523–1531, 2019.
28. Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. The fair guiding principles for scientific data management and stewardship. *Scientific data*, 3(1):1–9, 2016.
29. Dylan J Walsh, Weizhong Zou, Ludwig Schneider, Reid Mello, Michael E Deagen, Joshua Mysona, Tzyy-Shyang Lin, Juan J de Pablo, Klavs F Jensen, Debra J Audus, et al. Community resource for innovation in polymer technology (cript): a scalable polymer material data structure, 2023.
30. Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin A. Persson. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1), July 2013. ISSN 2166-532X. . URL `http://dx.doi.org/10.1063/1.4812323`.
31. Yukun Wu, Aikaterini Vriza, Doga Ozgulbas, Rafael Vescovi, Jianing Zhou, Zhiyang Wang, Shiyu Hu, Yuepeng Zhang, Qiaomu Yang, Anna Österholm, and et al. Autonomous synthesis and inverse design of electronic polymers with high efficiency and accuracy. *ChemRxiv*, 2025. .
32. Benjamin J. Blaiszik. Awesome matchem datasets. URL `https://github.com/blaiszik/awesome-matchem-datasets`.
33. Xue Jiang, Weiren Wang, Shaohan Tian, Hao Wang, Turab Lookman, and Yanjing Su. Applications of natural language processing and large language models in materials discovery. *npj Computational Materials*, 11(1):79, 2025.
34. Maciej P. Polak and Dane Morgan. Extracting accurate materials data from research papers with conversational language models and prompt engineering. *Nature Communications*, 15(1), February 2024. ISSN 2041-1723. . URL `http://dx.doi.org/10.1038/s41467-024-45914-8`.

35. Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. Survey of hallucination in natural language generation. *ACM computing surveys*, 55(12):1–38, 2023.
36. Anthony Yu-Tung Wang, Ryan J Murdock, Steven K Kauwe, Anton O Oliynyk, Aleksander Gurlo, Jakoah Brgoch, Kristin A Persson, and Taylor D Sparks. Machine learning for materials scientists: an introductory guide toward best practices. *Chemistry of Materials*, 32(12):4954–4965, 2020.
37. Maxwell L. Hutchinson, Erin Antono, Brenna M. Gibbons, Sean Paradiso, Julia Ling, and Bryce Meredig. Overcoming data scarcity with transfer learning, 2017. URL `https://arxiv.org/abs/1711.05099`.
38. Jiaqi Yang, Panayotis Manganaris, and Arun Mannodi-Kanakkithodi. Discovering novel halide perovskite alloys using multi-fidelity machine learning and genetic algorithm. *The Journal of Chemical Physics*, 160(6), February 2024. ISSN 1089-7690. . URL `http://dx.doi.org/10.1063/5.0182543`.
39. Hironao Yamada, Chang Liu, Stephen Wu, Yukinori Koyama, Shenghong Ju, Junichiro Shiomi, Junko Morikawa, and Ryo Yoshida. Predicting materials properties with little data using shotgun transfer learning. *ACS Central Science*, 5(10):1717–1730, September 2019. ISSN 2374-7951. . URL `http://dx.doi.org/10.1021/acscentsci.9b00804`.
40. Santisudha Panigrahi, Anuja Nanda, and Tripti Swarnkar. A survey on transfer learning. In *Intelligent and Cloud Computing: Proceedings of ICICC 2019, Volume 1*, pages 781–789. Springer, 2020.
41. Sandipp Krishnan Ravi, Yigitcan Comlek, Arjun Pathak, Vipul Gupta, Rajnikant Umretiya, Andrew Hoffman, Ghanshyam Pilania, Piyush Pandita, Sayan Ghosh, Nathaniel Mckeever, Wei Chen, and Liping Wang. Interpretable multi-source data fusion through latent variable gaussian process. *Engineering Applications of Artificial Intelligence*, 145:110033, April 2025. ISSN 0952-1976. . URL `http://dx.doi.org/10.1016/j.engappai.2025.110033`.
42. A. G. Kusne, D. Keller, A. Anderson, A. Zaban, and I. Takeuchi. High-throughput determination of structural phase diagram and constituent phases using GRENDEL. *Nanotechnology*, 26 (44), 2015. ISSN 13616528. .
43. A. Gilad Kusne, Heshan Yu, Changming Wu, Huairuo Zhang, Jason Hattrick-Simpers, Brian DeCost, Suchismita Sarker, Corey Oses, Cormac Toher, Stefano Curtarolo, Albert V. Davydov, Ritesh Agarwal, Leonid A. Bendersky, Mo Li, Apurva Mehta, and Ichiro Takeuchi. On-the-fly closed-loop materials discovery via Bayesian active learning. *Nature Communications*, 11(1): 1–11, 2020. ISSN 20411723. .
44. A. Gilad Kusne, Austin McDannald, and Brian DeCost. Learning material synthesis–process–structure–property relationship by data fusion: Bayesian co-regionalization N-dimensional piecewise function learning. *Digital Discovery*, 3(11):2211–2225, 2024. ISSN 2635-098X. . URL `http://arxiv.org/abs/2311.06228https://xlink.rsc.org/?DOI=D4DD00048J`.
45. Austin McDannald, Matthias Frontzek, Andrei T Savici, Mathieu Doucet, Efrain E Rodriguez, Kate Meuse, Jessica Opsahl-Ong, Daniel Samarov, Ichiro Takeuchi, William Ratcliff, and A Gilad Kusne. On-the-fly autonomous control of neutron diffraction via physics-informed bayesian active learning. *Appl. Phys. Rev.*, 9(2):021408, June 2022.
46. Izumi Takahara, Kiyou Shibata, and Teruyasu Mizoguchi. Generative Inverse Design of Crystal Structures via Diffusion Models with Transformers. 2024. URL `http://arxiv.org/abs/2406.09263`.
47. Kamal Choudhary and Brian DeCost. Atomistic line graph neural network for improved materials property predictions. *npj Computational Materials*, 2021. . URL `https://doi.org/10.1038/s41524-021-00650-1`.
48. Runze Zhang, Robert Black, Debashish Sur, Parisa Karimi, Kangming Li, Brian DeCost, John R. Scully, and Jason Hattrick-Simpers. Editors' Choice—AutoEIS: Automated Bayesian Model Selection and Analysis for Electrochemical Impedance Spectroscopy. *Journal of The*

Electrochemical Society, 170(8):086502, 2023. ISSN 0013-4651. .

49. Di Chen, Yiwei Bai, Sebastian Ament, Wenting Zhao, Dan Guevarra, Lan Zhou, Bart Selman, R. Bruce van Dover, John M. Gregoire, and Carla P. Gomes. Automating crystal-structure phase mapping by combining deep learning with constraint reasoning. *Nature Machine Intelligence*, 3(9):812–822, 2021. ISSN 25225839. .
50. Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4765–4774. Curran Associates, Inc., 2017. URL `http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf`.
51. Kangming Li, Brian DeCost, Kamal Choudhary, Michael Greenwood, and Jason Hattrick-Simpers. A critical examination of robustness and generalizability of machine learning prediction of materials properties. *npj Computational Materials*, 9(1), April 2023. ISSN 2057-3960. .
52. Kangming Li, Andre Niyongabo Rubungo, Xiangyun Lei, Daniel Persaud, Kamal Choudhary, Brian DeCost, Adji Bousso Dieng, and Jason Hattrick-Simpers. Probing out-of-distribution generalization in machine learning for materials. *Communications Materials*, 6(1), January 2025. ISSN 2662-4443. .
53. Yuxuan Wang and Ross D. King. Extrapolation is not the same as interpolation. *Machine Learning*, 113(10):8205–8232, 2024. ISSN 15730565. . URL `https://doi.org/10.1007/s10994-024-06591-2`.
54. Mert Yuksekgonul, Linjun Zhang, James Zou, and Carlos Ernesto Guestrin. Beyond Confidence: Reliable Models Should Also Consider Atypicality. *Advances in Neural Information Processing Systems*, 36(NeurIPS):1–34, 2023. ISSN 10495258.
55. Peter I. Frazier, Warren B. Powell, and Savas Dayanik. A knowledge-gradient policy for sequential information collection. *SIAM Journal on Control and Optimization*, 47(5):2410–2439, 2008. . URL `https://doi.org/10.1137/070693424`.
56. Maximilian Balandat, Brian Karrer, Daniel R. Jiang, Samuel Daulton, Benjamin Letham, Andrew Gordon Wilson, and Eytan Bakshy. Botorch: a framework for efficient monte-carlo bayesian optimization. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.
57. Xiaojin Zhu, John Lafferty, and Zoubin Ghahramani. Combining active learning and semi-supervised learning using gaussian fields and harmonic functions. In *ICML 2003 workshop on the continuum from labeled to unlabeled data in machine learning and data mining*, volume 3, pages 58–65, 2003.
58. Biswajit Paria, Kirthevasan Kandasamy, and Barnabás Póczos. A flexible framework for multi-objective bayesian optimization using random scalarizations. In Ryan P. Adams and Vibhav Gogate, editors, *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*, volume 115 of *Proceedings of Machine Learning Research*, pages 766–776. PMLR, 22–25 Jul 2020. URL `https://proceedings.mlr.press/v115/paria20a.html`.
59. Yicun Hua, Qiqi Liu, Kuangrong Hao, and Yaochu Jin. A survey of evolutionary algorithms for multi-objective optimization problems with irregular pareto fronts. *IEEE/CAA Journal of Automatica Sinica*, 8(2):303–318, 2021. .
60. Miqing Li, Shengxiang Yang, and Xiaohui Liu. Shift-based density estimation for pareto-based algorithms in many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 18(3):348–365, 2014. .
61. Samuel Daulton, Maximilian Balandat, and Eytan Bakshy. Differentiable expected hypervolume improvement for parallel multi-objective bayesian optimization. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.
62. Shinya Suzuki, Shion Takeno, Tomoyuki Tamura, Kazuki Shitara, and Masayuki Karasuyama.

Multi-objective Bayesian optimization using pareto-frontier entropy. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 9279–9288. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/suzuki20a.html.

63. Logan Saar, Haotong Liang, Alex Wang, Austin McDannald, Efrain Rodriguez, Ichiro Takeuchi, and A. Gilad Kusne. The LEGOLAS Kit: A low-cost robot science kit for education with symbolic regression for hypothesis discovery and validation. *MRS Bulletin*, 47(9):881–885, 2022. ISSN 08837694. . URL https://doi.org/10.1557/s43577-022-00430-2.
64. David P. Hoogerheide and Frank Heinrich. Autorefl: active learning in neutron reflectometry for fast data acquisition. *Journal of Applied Crystallography*, 57(4):1192–1204, July 2024. ISSN 1600-5767. . URL http://dx.doi.org/10.1107/S1600576724006447.
65. Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. OpenAI Gym. pages 1–4, jun 2016. URL http://arxiv.org/abs/1606.01540.
66. Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. The MIT Press, second edition, 2018. URL http://incompleteideas.net/book/the-book-2nd.html.
67. Sergei V Kalinin, Yongtao Liu, Arpan Biswas, Gerd Duscher, Utkarsh Pratiush, Kevin Roccapriore, Maxim Ziatdinov, and Rama Vasudevan. Human-in-the-loop: The future of machine learning in automated electron microscopy. *Microscopy today*, 32(1):35–41, 2024.
68. Felix Adams, Austin McDannald, Ichiro Takeuchi, and A Gilad Kusne. Human-in-the-loop for bayesian autonomous materials phase mapping. *Matter*, 7(2):697–709, 2024.
69. Marco Lombardi, Francesco Pascale, and Domenico Santaniello. Internet of things: A general overview between architectures, protocols and applications. *Information*, 12(2):87, 2021.
70. Howie Joress, Brian DeCost, Najlaa Hassan, Trevor M Braun, Justin M Gorham, and Jason Hattrick-Simpers. Development of an automated millifluidic platform and data-analysis pipeline for rapid electrochemical corrosion measurements: a ph study on zn-ni. *Electrochimica Acta*, 428:140866, 2022.
71. Howie Joress, Martin L Green, Ichiro Takeuchi, and Jason R Hattrick-Simpers. Applications of high throughput (combinatorial) methodologies to electronic, magnetic, structural, and energy-related materials. In *Encyclopedia of Materials: Metals and Alloys*, pages 353–371. Elsevier, 2022.
72. Daniel Jacobson, Greg Brail, and Dan Woods. *APIs: A strategy guide*. " O'Reilly Media, Inc.", 2012.
73. Marcus M Noack, Kevin G Yager, Masafumi Fukuto, Gregory S Doerk, Ruipeng Li, and James A Sethian. A kriging-based approach to autonomous experimentation with applications to x-ray scattering. *Sci. Rep.*, 9(1):11809, August 2019.
74. Mario Teixeira Parente, Georg Brandl, Christian Franz, Uwe Stuhr, Marina Ganeva, and Astrid Schneidewind. Active learning-assisted neutron spectroscopy with log-gaussian processes. *Nat. Commun.*, 14(1):2246, April 2023.
75. Marcus M Noack, Petrus H Zwart, Daniela M Ushizima, Masafumi Fukuto, Kevin G Yager, Katherine C Elbert, Christopher B Murray, Aaron Stein, Gregory S Doerk, Esther H R Tsai, Ruipeng Li, Guillaume Freychet, Mikhail Zhernenkov, Hoi-Ying N Holman, Steven Lee, Liang Chen, Eli Rotenberg, Tobias Weber, Yannick Le Goc, Martin Boehm, Paul Steffens, Paolo Mutti, and James A Sethian. Gaussian processes for autonomous data acquisition at large-scale synchrotron and neutron facilities. *Nat. Rev. Phys.*, 3(10):685–697, July 2021.
76. Runhai Ouyang, Stefano Curtarolo, Emre Ahmetcik, Matthias Scheffler, and Luca M. Ghiringhelli. Sisso: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Physical Review Materials*, 2(8), August 2018. ISSN 2475-9953. . URL http://dx.doi.org/10.1103/PhysRevMaterials.2.083802.
77. Alberto Hernandez, Adarsh Balasubramanian, Fenglin Yuan, Simon A. M. Mason, and Tim Mueller. Fast, accurate, and transferable many-body interatomic potentials by symbolic re-

gression. *npj Computational Materials*, 5(1), November 2019. ISSN 2057-3960. . URL http://dx.doi.org/10.1038/s41524-019-0249-1.

78. Zachary Trautt, Austin McDannald, Brian DeCost, Howard Joress, A. Gilad Kusne, Francesca Tavazza, and Tom Blanton. Workshop report on autonomous methodologies for accelerating x-ray measurements, 2024-11-05 05:11:00 2024. URL https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=958560.
79. Leslie Ching Ow Tiong, Hyuk Jun Yoo, Nayeon Kim, Chansoo Kim, Kwan-Young Lee, Sang Soo Han, and Donghun Kim. Machine vision-based detections of transparent chemical vessels toward the safe automation of material synthesis. *npj Computational Materials*, 10(1):42, 2024.
80. Sagi Eppel, Haoping Xu, Mor Bismuth, and Alan Aspuru-Guzik. Computer vision for recognition of materials and vessels in chemistry lab settings and the vector-labpics data set. *ACS central science*, 6(10):1743–1752, 2020.
81. Howie Joress, Zachary Trautt, Austin McDannald, Brian DeCost, A. Gilad Kusne, and Francesca Tavazza. Driving u.s. innovation in materials and manufacturing using ai and autonomous labs, 2024-08-14 04:08:00 2024. URL https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=958246.
82. Kamal Choudhary, Brian DeCost, Chi Chen, Anubhav Jain, Francesca Tavazza, Ryan Cohn, Cheol Woo Park, Alok Choudhary, Ankit Agrawal, Simon JL Billinge, et al. Recent advances and applications of deep learning methods in materials science. *npj Computational Materials*, 8(1):59, 2022.
83. Kamal Choudhary, Brian DeCost, and Francesca Tavazza. Machine learning with force-field-inspired descriptors for materials: Fast screening and mapping energy landscape. *Physical review materials*, 2(8):083801, 2018.
84. Logan Ward, Alexander Dunn, Alireza Faghaninia, Nils ER Zimmermann, Saurabh Bajaj, Qi Wang, Joseph Montoya, Jiming Chen, Kyle Bystrom, Maxwell Dylla, et al. Matminer: An open source toolkit for materials data mining. *Computational Materials Science*, 152:60–69, 2018.
85. Kevin Maik Jablonka, Qianxiang Ai, Alexander Al-Feghali, Shruti Badhwar, Joshua D Bocarsly, Andres M Bran, Stefan Bringuier, L Catherine Brinson, Kamal Choudhary, Defne Circi, et al. 14 examples of how llms can transform materials science and chemistry: a reflection on a large language model hackathon. *Digital discovery*, 2(5):1233–1250, 2023.
86. Kamal Choudhary. Atomgpt: Atomistic generative pretrained transformer for forward and inverse materials design. *The Journal of Physical Chemistry Letters*, 15(27):6909–6917, 2024.
87. Kamal Choudhary, Brian DeCost, Lily Major, Keith Butler, Jeyan Thiyagalingam, and Francesca Tavazza. Unified graph neural network force-field for the periodic table: solid state applications. *Digital Discovery*, 2(2):346–355, 2023.
88. Han Yang, Chenxi Hu, Yichi Zhou, Xixian Liu, Yu Shi, Jielan Li, Guanzhi Li, Zekun Chen, Shuizhou Chen, Claudio Zeni, et al. Mattersim: A deep learning atomistic model across elements, temperatures and pressures. *arXiv preprint arXiv:2405.04967*, 2024.
89. Kamal Choudhary. Diffractgpt: Atomic structure determination from x-ray diffraction patterns using a generative pretrained transformer. *The Journal of Physical Chemistry Letters*, 16(8): 2110–2119, 2025.
90. Kamal Choudhary. Microscopygpt: Generating atomic-structure captions from microscopy images of 2d materials with vision-language transformers. *The Journal of Physical Chemistry Letters*, 0(0):7028–7035, 0. .
91. Francesca Tavazza, Brian DeCost, and Kamal Choudhary. Uncertainty prediction for machine learning models of material properties. *ACS omega*, 6(48):32431–32440, 2021.
92. Kamal Choudhary, Daniel Wines, Kangming Li, Kevin F Garrity, Vishu Gupta, Aldo H Romero, Jaron T Krogel, Kayahan Saritas, Addis Fuhr, Panchapakesan Ganesh, et al. Jarvis-leaderboard: a large scale benchmark of materials design methods. *npj Computational Materials*, 10(1):93, 2024.

93. Alexander Dunn, Qi Wang, Alex Ganose, Daniel Dopp, and Anubhav Jain. Benchmarking materials property prediction methods: the matbench test set and automatminer reference algorithm. *npj Computational Materials*, 6(1):138, 2020.
94. Nawaf Alampara, Mara Schilling-Wilhelmi, and Kevin Maik Jablonka. Lessons from the trenches on evaluating machine-learning systems in materials science, 2025. URL `https://ww.arxiv.org/abs/2503.10837`.
95. Gary Tom, Stefan P. Schmid, Sterling G. Baird, Yang Cao, Kourosh Darvish, Han Hao, Stanley Lo, Sergio Pablo-García, Ella M. Rajaonson, Marta Skreta, Naruki Yoshikawa, Samantha Corapi, Gun Deniz Akkoc, Felix Strieth-Kalthoff, Martin Seifrid, and Alán Aspuru-Guzik. Self-driving laboratories for chemistry and materials science. *Chemical Reviews*, 124(16):9633–9732, August 2024. ISSN 1520-6890. .
96. Cosima Boswell-Koller, Benji Maruyama, James A. Warren, Charles Yang, Lisa E. Friedersdorf, Germano Iannacchione, and Richard Vaia. Accelerated Materials Experimentation Enabled by the Autonomous Materials Innovation Infrastructure (AMII) A Workshop Report. Technical report, Subcommittee on the MGI, Alexandria, VA, 2024. URL `https://www.mgi.gov/sites/mgi/files/MGI_Autonomous_Materials_Innovation_Infrastructure_Workshop_Report.pdf`.