Macromolecules

Editors' Choice

pubs.acs.org/Macromolecules

Perspective

Automation and Machine Learning for Accelerated Polymer Characterization and Development: Past, Potential, and a Path Forward

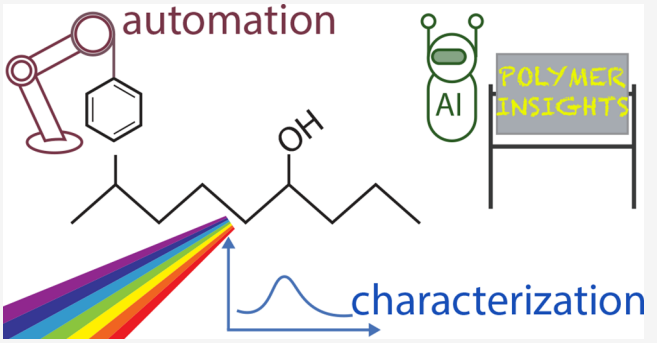
Peter A. Beaucage,* Duncan R. Sutherland, and Tyler B. Martin*

Cite This: https://doi.org/10.1021/acs.macromol.4c01410

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Automation and machine learning techniques are poised to dramatically accelerate the development of new materials while simultaneously increasing our understanding of the physics and chemistry that underlie the formation of such materials. In particular, the convergence of accessible machine learning tools, the availability of high-quality data, and the advent of accessible experimental automation platforms have led to a number of closed-loop autonomous experimentation platforms or "self-driving labs". Such platforms integrate robotic experimenters with AI-guided experiment planning to autonomously perform large numbers of experiments without human input. After briefly reviewing the state of the field and the broad classes of autonomous efforts, this perspective outlines several high-value focus areas for future ML-guided characterization efforts. Among many advantages, we expect that autonomous approaches will allow the systematic study of rare and nonequilibrium phenomena, provide dramatically greater measurement efficiency through targeting of cutting-edge, resource-intensive characterization, and enable a higher level of thinking and experimental planning for human investigators. Finally, we outline the principal barriers to realization of these advantages, including: (1) a lack of organizational structures and workforce development for the highly interdisciplinary programs needed; (2) funding and publication mechanisms that assign greater value to individual scientific results than foundational infrastructure development; and (3) a dearth of standards for open interchange of hardware, software, and data among the polymer community. We believe that we are in the early days of a once-in-a-generation shift in the way science is planned, executed, and evaluated, and we hope to provide a blueprint for the broader polymer community to take a leading role in this shift.



Downloaded via 195.252.220.180 on September 7, 2024 at 17:21:42 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

INTRODUCTION

Polymer science, and indeed materials science at large, is bearing witness to the convergence of several technological advancements: accessible, open-source libraries for complex machine-learning methods, a broad recognition of the intrinsic value of materials measurement data, and the emergence of low-cost, accessible robotic hardware. Increasingly, materials characterization is being conducted by automated robotic systems that can tirelessly synthesize and/or characterize materials with near-perfect precision and quantified uncertainty. Where physical and chemical models fall short, machine learning (ML) models are filling gaps with unparalleled speed and accuracy. The union of these approaches has been given many names (e.g., self-driving laboratory, autonomous platform, closed-loop experiments), but the results are common: faster discovery and more complete characterization of highly optimized materials. In this paradigm, the experimenter is granted a virtual assistant that guides the course of the experiment with varying degree of autonomy from the human scientist.

When reckoning with the impact of this paradigm shift, it is useful to consider the name "self-driving lab" as an analogy for the oft-used "self-driving car". The term "self-driving lab" may not be immediately intuitive to all audiences but the analogy to self-driving cars is an accessible way to convey the concept of an autonomous decision-making system. Self-driving laboratories (SDLs), like self-driving cars, envision a future where the scientist, or driver, can fully disengage from the task and relax while an AI system speeds them to their destination, scientific or physical, without input or guidance. Like present

Received: June 17, 2024
Revised: August 18, 2024
Accepted: August 26, 2024

ACS Publications

© XXXX American Chemical Society

realizations of self-driving cars, self-driving laboratories are better thought of as a hierarchy of guidance systems where even partial realizations are useful. A car on the "self-driving spectrum" might include features as basic as adaptive cruise control (SAE Level 1) or as advanced as autonomous lane guidance and route following with human oversight (SAE Level 3).[1] Autonomous functions can also be quite limited in scope and application and address only narrow aspects of the task at hand, such as automatic parallel parking. Certain components of an autonomous AI, such as traffic-aware route planning, might be routinely used to provide input to a human expert who remains in control without any hardware integration at all (e.g., running a map application on a mobile phone). The same vehicle might be operated in different modes at different times depending on the expertise and trust or comfort level of a human driver, for instance, engaging a route-following AI while on the highway but disabling automatic lane changes in dense traffic, disabling the autonomous system while executing a nuanced maneuver like allowing someone to merge in front, etc.

We consider each of these scenarios and operating modes to have analogies in the design and operation of a self-driving lab. For instance, an SDL might be operated with full closed-loop ML guidance of each step in an experiment and allowed to run unsupervised for a limited duration, such as overnight or over a weekend. It might also be used as essentially a "robotic lab technician", preparing and measuring samples from a user-defined list based on expert intuition. One might automate a particular task, like sampling of polymers from a tray of vials for chromatography, while still synthesizing the polymers and preparing samples by hand. An SDL might use different levels of engagement as the experiment progresses, such as seeding an experimental measurement campaign with expert choices of starting measurements before allowing the AI to run independently or disconnecting an AI that has found local extrema and inputting a set of manual conditions to encourage broader search. While the degree of autonomy leveraged by these different approaches varies, all represent increased efficiency in characterization of polymeric materials and complex fluids and meaningful progress toward more AI-assisted and fully autonomous experiments.

The adoption of self-driving laboratories faces challenges as the scientific world adapts to the AI revolution. While significant demonstrations and progress has been made toward this vision, fully autonomous self-driving laboratories remain more aspirational than realized for most researchers. As William Gibson aptly noted, "The future is already here — it's just not very evenly distributed." Developing and running self-driving laboratories is an unavoidably interdisciplinary challenge that requires domain knowledge across multiple fields. Material scientists understand the hypotheses, data processing, and experimental instruments necessary to answer their scientific questions, but often lack the proper coding or automation skills. Instrument building requires skills in mechanical and software engineering, which can present a high barrier to development of autonomous laboratories, especially if the engineering is not related to typical tasks within the instrument's scientific domain. While computer and data scientists excel at implementing state-of-the-art ML and AI methodologies, there remains a "productization gap" when transitioning these methods into useful real-world tools that physical scientists can rely upon. This challenge often stems from the differing incentives and focus areas between computer science research and the practical needs of domain scientists, rather than a fundamental limitation of CS methods or computer engineers. Further complicating these matters, the modern funding environment and scientific reward structure has struggled to find ways to support the development and dissemination of autonomous platforms. Promising demos and proofs of concept need to be transformed into reliable, user-friendly tools for the broader scientific community. Both the hardware and software associated with these tools need user manuals, tutorials, and ongoing support as bugs are found and features are added. Current academic incentives and funding mechanisms are often not well-aligned with the sustained effort required to bridge this gap, creating a significant barrier to the widespread adoption of these powerful capabilities.

In this perspective, we briefly review the state of the field of autonomous SDLs as applied to polymer science and engineering, outline our perspective on growth areas and vision for a future where polymer science is more efficient, edge cases are better understood, and results are more reliable.

PAST (AND PRESENT)

The use of automation and high-throughput techniques in sample preparation and characterization has been a mainstay of applied polymer science for decades.[2,3] The combinational and high-throughput science revolution that first began in the 2000s continues to demonstrate the utility of rapid materials screening and property measurement through customized sample synthesis and robotic automation techniques. These concepts have grown into fully realized robotic and microfluidic platforms that synthesize and characterize polymer materials with minimal interactions from human operators.[4,5] While these platforms have been of great utility, it is important to note that their designs are also often bespoke and inflexible, making it hard to adapt them to material challenges outside of their original design specifications.

These breakthroughs in automated experimentation were (and are) being driven by the needs of fast-paced industrial research. The nature of the studies needed to translate foundational breakthroughs to products necessitates the exploration of highly complex and constrained parameter landscapes at a pace beyond what manual operators can support. The growth of automation is further driven by simple labor economics: robots will work constantly without breaks at a cost inevitably lower than a human instrument operator, unlocking increased throughput. Beyond throughput, biosciences and routine chemical analytics have led the charge into automation for an increase in measurement reliability and traceability.

The rise of autonomous measurements began decades after the initial rise of automation. While there are examples of closed-loop experimentation in various fields,[6−9] the development of ARES represented a significant advancement in materials synthesis and characterization driven by an artificial intelligence (AI) engine.[10] Since then, a variety of other SDLs have been created, all attempting to leverage our past advances in automation in the new age of machine learning and data science. AI agents driving synthetic platforms are tuning polymer sequence, molar mass distribution, or simply the efficiency of polymer syntheses.[11−15] Other agents are focused on optimizing the properties of polymer materials or formulated mixtures.[14,16] All of these agents attempt to leverage the tireless speed of robotic platforms with machine-

https://doi.org/10.1021/acs.macromol.4c01410
Macromolecules XXXX, XXX, XXX−XXX

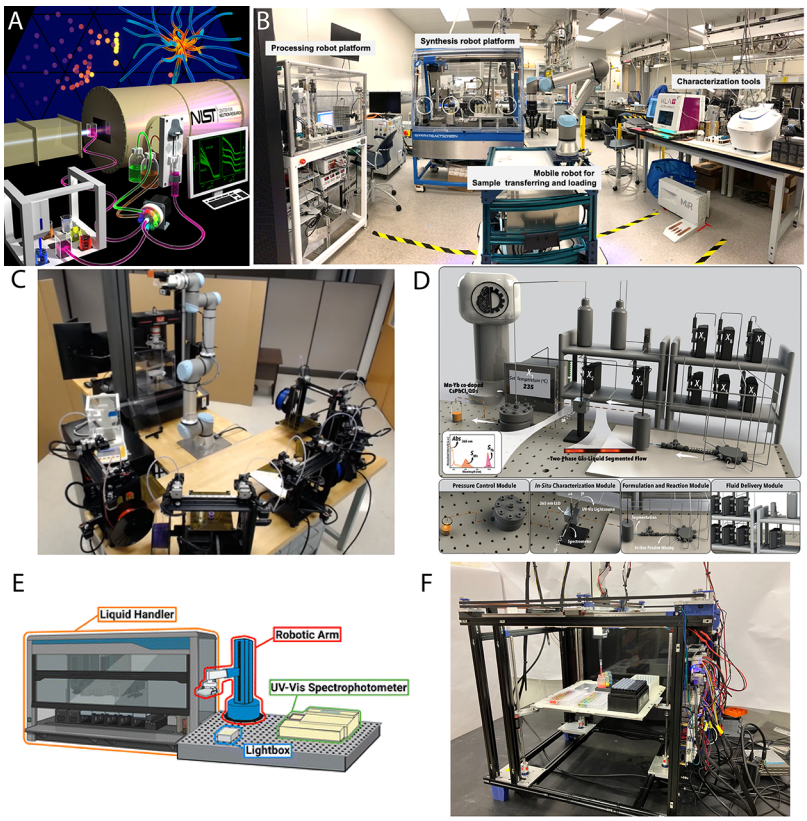


Figure 1. Examples of autonomous experiment platforms. (A) NIST Autonomous Formulation Laboratory, (B) Argonne PolyBot Platform,[17] (C) Bayesian Experimental Autonomous Researcher (BEAR),[18] (D) Smart Dope,[19] (E) PolyTron,[20] and (F) Jubilee open lab automation platform, used for sonochemical synthesis.[21] (B) Reproduced from ref 17. Copyright 2023 American Chemical Society. (C) Reproduced with permission from ref 18. Copyright 2020 The American Association for the Advancement of Science. (D) Reproduced with permission from ref 19. Copyright 2024 John Wiley and Sons. (E) Reproduced with permission from ref 20 under a Creative Commons Attribution 3.0 Unported License. (F) Reproduced with permission from ref 21 under a Creative Commons Attribution-NonCommercial 3.0 Unported License.

guided precision to help optimize and discover new polymer materials and formulations (Figure 1).

Even with these advancements, we find ourselves at a bifurcation of paths in self-driving laboratories. As the transformative potential of these techniques has become clear, large-scale investment has been pursued and a conversation created around large, national user facility-style laboratories. Some nascent programs of note include the PolyBot program at Argonne and the Autonomous Formulation Laboratory at NIST.[17,22] Such facilities are by no means limited to government, with autonomous user programs in active development in the University of Toronto-led Acceleration Consortium and in other academic settings.[23−25] All of these platforms center around investments on the scale between hundreds of thousands and millions of US dollars. They offer enormous potential for accessibility but little potential for ownership, as few laboratories or funding agencies have programs dedicated to investing such sums in infrastructure (see below), and those that exist are highly oversubscribed. Many are on funding clocks with a proof of concept or initial result ending the continued operation. Furthermore, many of the highest-performing platforms are constructed by contractors with highly bespoke components and limited potential for reconfiguration. This means that once the initial objective of the study is complete the infrastructure investment is essentially single-use unless care is taken in designing a general system, sometimes as a trade-off with performance for the designed task.

In parallel, several PIs have focused on developing open-source laboratory automation using small-scale instrumentation typically under $50k USD.[26] Generally, these unique platforms are designed and constructed by scientists whose motivations lie in domain science, such as polymer or material science, rather than pure AI or robotics groups. Despite dramatic differences in investment, the results have been no less impactful than those from national facility-scale laboratories; on the contrary, the accessibility of the instrumentation has made high-risk/high-reward experiments and extensive iteration possible. Notable examples of this paradigm include the Bayesian Experimental Autonomous Researcher platform for mechanical testing,[18] a sonochemical synthesis platform,[21] and the PolyTron photochemical synthesis platform.[20] There are also compelling arguments that instrument engineering and software development are best left to, or at least conducted under the close guidance of, professionals. However, a growing societal shift toward a "maker mindset" has made software development and digital fabrication technologies part of the curricula at younger and younger stages, and recent advances in stereolithography (SLA) additive manufacturing and high-quality measurement electronics have made these techniques both useful for scientific work and within reach of small-to-medium-scale budgets and timeframes.

https://doi.org/10.1021/acs.macromol.4c01410
Macromolecules XXXX, XXX, XXX−XXX

A parallel advance to the march of automated experimentation has occurred in the area of data-driven science and machine learning for polymers. We note that this area has recently been reviewed very comprehensively,[27] so here we highlight only a few efforts of particular relevance to characterization of soft materials. Broadly, these advancements represent novel platforms and approaches for automated data analysis and facile methodologies to represent and store polymer chemistry information, such as sequence and chain length, in computer databases. ML tools, such as CREASE, are being used provide better fits and extract more information from neutron and X-ray scattering measurements.[28] There have also been works using Bayesian and information theory methods to optimize neutron reflectivity measurements.[29,30] DeepStruc predicts crystal structures from pair-distribution function measurements, and while not originally applied to soft-materials, it demonstrates several powerful concepts (e.g., graph conditional variational autoencoders) that are immediately translated to our field.[31] Large, community-supported polymer description schemes and data warehouses, such as BigSMILES and CRIPT, have been developed and are already being used.[32,33]

Finally, a promising use case of materials ML is a set of methods we have come to colloquially refer to as "measurement transmutation" (also commonly known as proxy measurement). Put more rigorously, this is the inference of a high-cost measurement (such as X-ray or neutron scattering) from a known and controlled application space and a low-cost measurement (such as optical spectroscopy) and machine learning techniques rather than analytical data evaluation. PairVAE uses specially trained autoencoders to jointly encode electron microscopy and scattering measurements into a common latent space.[34] Using this approach, one measurement can be transmuted into the other, allowing researchers to leverage the real-space interpretability of microscopy and the global structural averaging of scattering. Another recent example of this paradigm is our recent work which demonstrated correlation between polymer architectural parameters authoritatively measured via size-exclusion chromatography and near-infrared (NIR) spectra.[35] While NIR is in principle sensitive to only local atomic order, machine learning techniques such as functional principal component analysis (fPCA) were able to recover the polymer conformation from subtle shifts in peak features, even though NIR is in principle only sensitive to local atomic order. While such methods are inherently heuristic and may never be appropriate for a research environment where new materials lie outside the training data set, they may open up new avenues for physics-based polymer characterization at scales previously thought impossible, such as postconsumer plastics recycling at the 100+ ton per day scale.

POTENTIAL

Automation Enables the Routine, Rigorous Study of Statistics. It can be argued that the central challenge in polymer science is the description and engineering of statistical ensembles of molecules. While some polymer characterization explicitly measures distributions (e.g., chromatography), many techniques make single-point measurements of single samples. Likely due to their relative expense in time and instrumentation, structural measurements such as microscopy and scattering frequently fall into this category. One simple, yet key advance enabled by measurement automation is an increase in the number of measurements that support a given conclusion, allowing uncertainty measurement across multiple samples as is routinely done in the life sciences. Such techniques will allow the quantification of parameters important for applied and theoretical studies alike.

This increase in data volume, beyond estimation of uncertainty, is coupled with the intrinsic reproducibility of automated sample preparation and measurement. The user-to-user and site-to-site variability of synthetic, formulation, and other experimental procedures is significant, yet poorly documented. Frequently, difficult-to-control environmental parameters such as barometric pressure, elevation, humidity, and ambient dust as well as minute equipment or supplier variations can play a key role in synthesis or characterization. Processing parameters such as pipetting speed and technique, stirring/shaking, or other flow-induced shear phenomena can lead to differences in material microstructure and properties.[36] In addition, much attention has been paid in recent years to polymer systems exhibiting pathway complexity, perhaps the most notable among them the existence of low-symmetry cubic phases such as the Frank-Kasper sigma phase in block copolymer melts.[37] In such systems, pioneering work by Dorfman and Bates showed a complex dependence of micelle size and tendency to form a given phase on the thermal pathway.[38] Human-driven, bespoke experimentation took years to find this significant result, but rigorous automated approaches that perform exactly the same actions, exactly the same way, every time can reduce such irreproducibility. How many people observed such low-symmetry sphere phases but assumed it was a result of contamination or preparation error?

The convergence of data volume and increased reproducibility offers significant potential to make edge phenomena dramatically more well-defined, to the point that they are worthy of deeper investigations. Most experimental investigations, especially looking near transitions of phase or property, have "odd samples" that are frequently not reported or repeated unless they have notable properties. The typical assumption is that the sample is either (a) representative of a human error in preparation or (b) the result of a hopelessly far-from-equilibrium state and not worthy of further investigation. Automated sample preparation promises to increase the reproducibility of such samples and at minimum will allow the close exploration of regions around such samples (i.e, with only slightly different processing parameters) and the estimation of the frequency of occurrence of the behavior. This presents a significant, yet to our knowledge minimally explored, route toward understanding the far-from-equilibrium behavior of polymers.

While automated robotic systems have the potential to improve reproducibility, we must also acknowledge that they also have the capacity to amplify mistakes in experimental design or erroneous AI/ML model assumptions at scales beyond what individual researchers are capable of. An assumption about the nature of a material or measurement can lead to an unattended autonomous platform wasting resources and time on an unattainable goal. The repeatability of autonomous platforms makes it easier to detect, and sometimes correct, systematic errors like this, but human researchers must be engaged to identify and correct the error. As discussed later, this is one reason why the most effective autonomous platforms will be collaborative rather than completely independent.

https://doi.org/10.1021/acs.macromol.4c01410
Macromolecules XXXX, XXX, XXX−XXX

Automation and Machine Learning Allow Targeting of Rare Measurements. Beyond the benefits of reproducibility and targeted investigation of edge cases, autonomous science approaches guided by machine learning offer the potential to dramatically increase the efficacy of cost- or time-intensive measurements through multimodal precharacterization and sample targeting. As representative examples of such experiments, consider neutron dynamics techniques such as QENS, surface rheology, isotope-labeled scattering and NMR studies, etc. In such measurements, considering more than a handful of individual samples would be a heroic endeavor representative of massive resources that can be allocated to a single study only in exceptional cases. The typical experimental progression, therefore, often entails prior understanding of an experimental/sample design space (e.g., from past literature studies, proposed physical mechanisms, and prior characterization conducted by the investigators). While this is a reasonable approach given the present state of the art, it is intrinsically qualitative and vulnerable to biases: a proposed model that links multiple data types can only rarely provide statistical foundation for those links.

Data fusion approaches built around statistical models such as Gaussian processes and machine learning tools, together with correlation-based data inference or "transmutation" tools, offer the potential to make sample selection for intensive measurements the product of cohesive, statistical estimations of the behavior of the sample and the impact of processing parameters on the sample. Uncertainty evaluation approaches provide a rigorous underpinning for what information the proposed intensive study would provide and how it might change or illuminate current understanding. We expect that such prior studies might become a routine part of proposals for beamtime at user facilities or other gates for access; the justification provided to the community for why one study over another should be conducted may involve statistical proof of the importance of samples in reducing uncertainty for a given task.

Autonomous and Human-Machine Teaming Is a Force Multiplier. Most science students and graduates experience some variant of the following story:

> *Working late into the night over many weeks, the scientist runs the 10th (or 100th or 1000th) iteration of a measurement/synthesis/simulation. Quick looks at the intermediate data during collection seem fine, but when they finally sit down to look at the full corpus of results, they discover that the instrument was miscalibrated, or the starting materials were contaminated, or their input files contained an error.*

While it is easy to view this as a lesson in being detail-oriented and careful in execution, perhaps the lesson is that the burden on the experimenter should be reduced. The truth is that carrying out an experiment is often a rote, nonintellectual task, and students and graduates should focus more on the analysis of past experiments and the designing of the new ones. Automated and autonomous platforms promise to revolutionize the way we do science by freeing up the experimenter to spend more time thinking about their science. Once developed, these platforms can tirelessly conduct experiments with near perfect precision at speeds that are impossible for humans to match. It is important to note that these speeds go beyond efficiency and can enable science that was previously impossible. The doubling or tripling of data set sizes that were previously constrained by beamtime or resource limitations allows for statistical and modeling analyses that are impossible on a singular measurement or small data set.

With the above benefits of autonomy outlined, we must acknowledge the value of hands-on experimental work in giving researchers the necessary foundation to design experiments and troubleshoot problems. A.I. agents cannot yet design whole materials studies or leverage physical principles to design a measurement apparatus. However, we also recognize that as technology evolves, so too must our approaches to education and training. Just as we no longer require scientists to master slide rules or memorize extensive tables of integrals due to the advent of computational tools, the role of hands-on experimentation may evolve. The key is to ensure that researchers develop a deep understanding of experimental principles and critical thinking skills, which can be achieved through a combination of hands-on work and engagement with automated systems.

This revolution in the way we conduct measurements, or, indeed, experiments at large, should not be viewed as replacing the human scientist with a robotic one but rather giving the human scientist a robotic collaborator. Human-machine teaming allows scientists to balance scientific intuition against hard-statistical and model-based analysis in real-time. This is the vision behind automakers releasing SAE Level 3 vehicles. While these self-driving vehicles require human attention and intervention in complex scenarios, they still provide significant utility to drivers and have the potential to reduce collisions. Analogously, we must not miss out on the benefits of a partially self-driving lab which collaborates with a human operator. A "SAE Level 3" self-driving lab is much easier to implement than a "SAE Level 5" and can still provide speedups and optimized materials. While Level 5 can still be an ultimate goal, Level 3 systems exist today and can be upgraded when the improved agents are ready for deployment. Furthermore, studies have shown that the human-machine teaming paradigm (also called human-in-the-loop) allows autonomous platforms to leverage human intuition and scientific knowledge that would be challenging to incorporate directly into the AI model.[39]

PATH FORWARD

With the past, present, and promise of self-driving laboratories outlined, the question remains of how to go from the former to the latter. Despite great demonstrations of the success and utility of autonomous approaches in the polymer science community and beyond, significant challenges remain. These challenges fall into several broad categories: education and workforce development, open and accessible hardware and software, improved cohesion between theory and data, outdated funding and performance models, and the need for standard- and guideline-focused community groups. In the following section, we will outline the nature of these challenges and provide paths and opportunities for mitigating them.

Training and Workforce Development Needs for the Autonomous Future. As discussed above, creating and operating a self-driving lab is an unavoidably interdisciplinary activity. While teams can and have come together to design these systems, researchers with appropriate multidisciplinary training will be essential to the broad adoption of autonomous science. Even partial fluency in two or more of the defining categories (i.e., hardware/robotic design, software development, machine learning, domain science) will improve translation and communication within a team. While there are short courses and webinars that teach either basic scientific

https://doi.org/10.1021/acs.macromol.4c01410
Macromolecules XXXX, XXX, XXX−XXX

programming or machine learning, these skills must be developed beyond what is possible in several days of lecture. Traditional educational programs in polymer science and related fields should pair their training, at the undergraduate and graduate levels, with training in hardware/robotics, software development, data science and/or machine learning. It bears noting that the National Science Foundation has recently funded several NRT (NSF Research Traineeship) programs to develop curricula that combine materials science and machine-learning training, indicating a growing recognition of this need.[40] These skills will be translatable far beyond the scope of this article as the world progresses further into the age of robotics and machine intelligence. At a minimum, young scientists must be trained to have basic proficiency in scientific software development beyond basic programming, including such software engineering concepts as version control, unit testing, continuous integration, semantic versioning, and general collaborative development practices. And, of course, the increased training in these new skill areas should not diminish the training in traditional polymer physics and chemistry. It is the combination of skills in automation, autonomy, and domain science that are needed to bring about a true change in the way we conduct scienctific studies.

Open Standards for Hardware, Software, and Data Interchange. Even when designed using off-the-shelf commercial components, the cost, time, and labor involved in developing autonomous platforms is often large and puts them out of reach for many research groups and organizations. Furthermore, this investment can be hard to justify if the platform might only be useful for a single study. The key to widespread adoption of autonomous science lies in the development of modular, open, and community-driven designs for self-driving laboratories. These hardware designs and software libraries should live beyond the supplementary files of journal articles and be released on platforms, such as GitHub or GitLab, where others can contribute to or expand on them. Software and hardware should be designed as combinations of single-concept, interchangeable components so that those components are more easily and likely to be reused.[41] We should support the development of software tools and interfaces, such as HuggingFace.co or MLExchange.[42,43] These tools democratize ML model development by providing a platform where users can upload, design, train, and share models. In the case of the latter,[42] MLExchange is designed to address some of the unique needs of the scientific community such as working with small data sets or mitigating the effects of measurement artifacts.

While the promise of open hardware and software is clear, there is a need for sustainable funding models and community engagement to support such hardware and software libraries over the long-term. The issues of continued maintenance, documentation, and support for bug fixes and upgrades must not be understated. Even for large, highly successful open scientific projects, there is often disproportionate effort put forward by small teams of unpaid and under-credited volunteers. For open science, and by extension autonomous science, to succeed, we need changes in the scientific culture around shared credit for the collaborative development of research tools.

Efforts to incentivize the shift of traditional commercial instruments from bespoke, closed, single-purpose devices to interchangeable, repairable, and programmable components will be particularly valuable; true open hardware projects have the double disadvantage that the time investment of fabrication is substantial and each update, unlike in software, has a marginal cost. Commercial systems designed for reconfigurability and reusability offer the potential to reduce that burden while maintaining many of the advantages of open hardware.

Somewhat separate from open hardware and software, efforts such as CRIPT that provide open data warehousing are a promising step toward generalized development of ML tools, though it is worth community consideration of where such databases should be held (individual universities, consortia, government agencies, nonprofits, etc.) given the complex legal and intellectual property considerations involved. More efforts in schema standardization with input from experienced experimental researchers is needed to make such platforms fully reflective of the diversity of polymer characterization data.[44]

Physics-Informed Machine Learning Modeling. The next generation of machine learning models and agents for self-driving platforms must move beyond the "black box" paradigm. Incorporation of physical and chemical models into machine-learning pipelines allows them to be more extrapolative and interpretable.[45] This incorporation can come in a variety of forms. Examples of this are difference or quotient learning, where models can learn the difference or ratio between a theoretical prediction and experimental data.[46] Separately, ML models can be trained to predict the parameters of physical or chemical models.[47] In this way, the predictions of the full pipeline will always be bounded by the scientific model. Physics-informed neural nets (PINNs) and graph neural networks use specific architectures and loss-function constructions to incorporate theory directly into the model.[48,49] While there are some promising works in using theory-informed machine learning outside of self-driving laboratories, these concepts have been much less applied in autonomous agents.

Beyond the incorporation of theory, models which produce uncertainty estimates, such as Gaussian processes (GPs), should also be preferred, although it is crucial to understand the limitations and details of the uncertainty estimation. For example, while GPs produce uncertainty estimates, these uncertainties may be unlike the true uncertainty of a prediction, e.g., a homoscedastic GP cannot properly describe the uncertainty of a heteroscedastic process, kernel design has a strong impact on the extrapolative and predictive power. Similarly, the uncertainty of variational GPs for classification is connected to the choice of "link function" and may not be truly reflective of the certainty in a class label. An alternative to using GP-based models, which do not perform well in certain cases such as with large data sets, is to consider statistical variants of models such as Bayesian neural networks.

20th Century Funding for 21st Century Infrastructure. The disruptive, generational shift in laboratories from manual to highly automated and data-driven infrastructure could be greatly accelerated by the development of new funding paradigms that support infrastructure and enable the education and development of students as intrinsically cross-disciplinary experts. Present models support scientific work, but only support the needed infrastructure through either equipment support on grants that primarily fund achievement of a scientific objective or through rare programs that directly fund the purchase of (mainly commercial) equipment. Few programs exist that directly fund the development of scientific infrastructure, which has traditionally been viewed as an

https://doi.org/10.1021/acs.macromol.4c01410
Macromolecules XXXX, XXX, XXX−XXX

overhead expense, yet in this moment of striking change this lack of appropriate funding risks creating "have" and "have not" laboratories. Similarly, while broad government support for education and training has enabled countless advances, insufficient support is available for development of new curricula, particularly new curricula focused on interdisciplinary needs. Such funding is essential for the full realization of the autonomous science paradigm shift. In particular, funding programs for the purchase of infrastructure should be enhanced in order to provide support for laboratory-scale investments of a general nature. In addition, such programs should also support the intensive development of new infrastructure, including through in-house engineering and development. Translation strategies must exist for such infrastructure to be shared, of course. Such funding is of limited effect if there are no students and postdocs with appropriate prior experience to be effective contributors to such projects; educational funding that would support the intensive work needed in development and retooling of curricula is also sorely needed.

In addition to funding for students and postdocs, there is a critical need for support of permanent nonprofessor technical staff. These roles are essential for maintaining and operating complex laboratory equipment over the long-term, which is crucial for the sustainability of academic SDL platforms. While automated systems can work continuously at potentially lower costs than human researchers, the maintenance and upkeep of these platforms (including their software, models, etc.) represent a significant ongoing cost and must be accounted for.

It is worth noting that a long-term shift to an autonomous future is somewhat misaligned with the incentives for career development of students and postdocs or for that matter even PI-level staff; by their nature, autonomous projects require effort from a variety of people with a variety of expertise, which intrinsically conflicts with the model of a, single first author of a manuscript. Taken *ad absurdum*, one could imagine autonomous systems leading to a dramatic upheaval of the academic research environment, though it is worth noting that industrial and government research programs frequently involve broad, diverse teams that share credit for accomplishments. In any case, whatever changes come about to the research system, it must be led by the demonstrated, overwhelming advantage of autonomy, a goal toward which many are making significant progress, yet one which remains for now largely unrealized. By realigning our research funding programs for the AI-driven 21st century, we have a chance to maximize the impact of a generational paradigm shift.

Open Hardware and Software Communities for Autonomous Polymer Science. For the promise of self-driving laboratories to be realized, a community of developers and users that focus on democratizing access to these tools must be created. The CASE group and the Acceleration Consortium are great examples of broad autonomous science-focused communities, but more are needed.[50,51] Other groups and communities should be formed that either focus on specific scientific domains (e.g., polymer science and engineering) or take a more standards development role. For the latter, such a group should develop open hardware and software guidelines that lower the barrier to accessing automated and autonomous experimentation solutions. These guidelines need not be prescriptive but should provide guideposts that increase the chances of success when developing a new self-driving lab. Even curated lists of open hardware and software would be helpful in jump-starting new projects. In particular, the community should agree on recommended practices when developing instrumentation and data APIs. This will not only increase the usefulness of smaller platforms but will make it easier for scientists to transfer their experiments to user-facilities if they share common or interoperable protocols. While they are likely too narrow for the polymer or larger autonomous community, efforts such as the "Sample Environment Communication Protocol (SECoP)" and PyLabRobot may serve as a starting template.[52,53]

CONCLUSION

Like society at large, science is in the midst of grappling with the staggering pace of advances in data science, AI, and machine learning. In times of abrupt change brought about by breakthrough discoveries, it is always difficult to separate lasting impacts from passing fads. On the basis of progress to date and reasonably expected growth, we believe that autonomous approaches have the potential to transform the way polymer science is done in a lasting and meaningful way: not only improving the efficiency of measurements but also enabling studies with new reliability and reproducibility and increasing our insights in subtle correlations of measurement data across many experiment types. However, this transformative potential demands changes of us as experimental practitioners, too: changes in the way we collaborate on projects, the way we develop the workforce of tomorrow, and the way we share software, hardware, and data. None of these challenges are insurmountable, and we believe that, as a result of the underlying foundational advances, the future of polymer science at large is brighter than it has been in the last half century.

AUTHOR INFORMATION

Corresponding Authors

Tyler B. Martin − *Material Science and Engineering Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States;* orcid.org/0000-0001-7253-6507; Email: tyler.martin@nist.gov

Peter A. Beaucage − *NIST Center for Neutron Research, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States;* orcid.org/0000-0002-2147-0728; Email: peter.beaucage@nist.gov

Author

Duncan R. Sutherland − *University of Colorado Boulder, Boulder, Colorado 80309, United States; Material Science and Engineering Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.macromol.4c01410

Notes

The authors declare no competing financial interest.

https://doi.org/10.1021/acs.macromol.4c01410
Macromolecules XXXX, XXX, XXX−XXX

Biographies

Dr. Peter Beaucage is a staff scientist at the NIST Center for Neutron Research and coleader of the Autonomous Formulation Laboratory project, which seeks to develop tools for the broad application of AI/ML in X-ray and neutron science in close collaboration with industry, government, and academic partners. His other interests include the development and application of resonant soft X-ray scattering (RSoXS), particularly for macromolecular solutions and biomolecules, and the application of autonomous and high-throughput scattering to materials challenges in energy, water, and climate. Peter's Ph.D. at Cornell University focused on the development of quantum metamaterials, using block copolymers to produce 3D mesostructures in superconductors.

Dr. Duncan Ross Sutherland is a postdoctoral fellow at the University of Colorado Boulder, where he works with the NIST Autonomous Formulation Laboratory program toward multimodal autonomous experimentation leveraging small angle scattering and spectroscopic techniques for reformulation challenges. His B.S. and M.S. in Chemistry from the University of Oregon in 2016 focused on solid state reaction chemistry and synthesis of metastable thin film chalcogenide materials. Afterward, he worked for Thermo Fisher Scientific on automation workflows for TEM preparation with the G4/EX 300 mm wafer dual beam FIB-SEM tools. His Ph.D. in Materials Science and Engineering at Cornell University focused on the design and implementation of a fully autonomous metal oxide thin film synthesis and characterization platform.

Dr. Tyler Martin is a staff member in the Materials Science and Engineering Division at NIST and a neutron beamline scientist for the *n*Soft consortium at the NIST Center for Neutron Research. Working closely with *n*Soft stakeholders, he leverages machine learning, molecular simulation, and liquid state theories to enhance neutron and X-ray scattering measurements of soft materials. Tyler coleads the Autonomous Formulation Laboratory program, which combines machine learning with automated measurement with the goal of accelerating formulation discovery and optimization. Tyler's Ph.D. at the University of Colorado focused on using simulation and theory to develop design rules for tailoring polymer nanocomposite morphology.

■ ACKNOWLEDGMENTS

We thank Dr. Chelsea Edwards of the NIST Center for Neutron Research and Dr. Christopher Stafford of the NIST Materials Science and Engineering Division for helpful comments and discussions. Certain commercial equipment, instruments, materials, organizations, or software are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

■ REFERENCES

(1) Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles; J3016_202104; SAE International, 2021.

(2) Meredith, C.; Smith, A. P.; Crosby, A. J.; Amis, E. J.; Karim, A. Combinatorial Methods for Polymer Science. In *Encyclopedia of Polymer Science and Technology*; Wiley, DOI: 10.1002/0471440264.pst069.

(3) Hoogenboom, R.; Meier, M. A. R.; Schubert, U. S. Combinatorial Methods, Automated Synthesis and High-Throughput Screening in Polymer Research: Past and Present. *Macromol. Rapid Commun.* **2003**, *24* (1), 15−32.

(4) Zaquen, N.; Rubens, M.; Corrigan, N.; Xu, J.; Zetterlund, P. B.; Boyer, C.; Junkers, T. Polymer Synthesis in Continuous Flow Reactors. *Prog. Polym. Sci.* **2020**, *107*, 101256.

(5) Tamasi, M.; Kosuri, S.; DiStefano, J.; Chapman, R.; Gormley, A. J. Automation of Controlled/Living Radical Polymerization. *Advanced Intelligent Systems* **2020**, *2* (2), 1900126.

(6) Winicov, H.; Schainbaum, J.; Buckley, J.; Longino, G.; Hill, J.; Berkoff, C. E. Chemical process optimization by computer — a self-directed chemical synthesis system. *Anal. Chim. Acta* **1978**, *103* (4), 469−476.

(7) MATSUDA, R.; ISHIBASHI, M.; TAKEDA, Y. Simplex Optimization of Reaction Conditions with an Automated System.

https://doi.org/10.1021/acs.macromol.4c01410
Macromolecules XXXX, XXX, XXX−XXX

CHEMICAL & PHARMACEUTICAL BULLETIN **1988**, *36* (9), 3512−3518.
(8) King, R. D.; Whelan, K. E.; Jones, F. M.; Reiser, P. G. K.; Bryant, C. H.; Muggleton, S. H.; Kell, D. B.; Oliver, S. G. Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature* **2004**, *427* (6971), 247−252.
(9) Chan, E. M.; Xu, C.; Mao, A. W.; Han, G.; Owen, J. S.; Cohen, B. E.; Milliron, D. J. Reproducible, High-Throughput Synthesis of Colloidal Nanocrystals for Optimization in Multidimensional Parameter Space. *Nano Lett.* **2010**, *10* (5), 1874−1885.
(10) Nikolaev, P.; Hooper, D.; Webber, F.; Rao, R.; Decker, K.; Krein, M.; Poleski, J.; Barto, R.; Maruyama, B. Autonomy in materials research: a case study in carbon nanotube growth. *npj Computational Materials* **2016**, *2* (1), 16031.
(11) Upadhya, R.; Kosuri, S.; Tamasi, M.; Meyer, T. A.; Atta, S.; Webb, M. A.; Gormley, A. J. Automation and data-driven design of polymer therapeutics. *Adv. Drug Delivery Rev.* **2021**, *171*, 1−28.
(12) Reis, M.; Gusev, F.; Taylor, N. G.; Chung, S. H.; Verber, M. D.; Lee, Y. Z.; Isayev, O.; Leibfarth, F. A. Machine-Learning-Guided Discovery of F-19 MRI Agents Enabled by Automated Copolymer Synthesis. *J. Am. Chem. Soc.* **2021**, *143* (42), 17677−17689.
(13) Tamasi, M. J.; Patel, R. A.; Borca, C. H.; Kosuri, S.; Mugnier, H.; Upadhya, R.; Murthy, N. S.; Webb, M. A.; Gormley, A. J. Machine Learning on a Robotic Platform for the Design of Polymer-Protein Hybrids. *Adv. Mater.* **2022**, *34*, 2201809.
(14) Knox, S. T.; Parkinson, S. J.; Wilding, C. Y. P.; Bourne, R. A.; Warren, N. J. Autonomous polymer synthesis delivered by multi-objective closed-loop optimization. *Polym. Chem.* **2022**, *13* (11), 1576−1585.
(15) Rubens, M.; Vrijsen, J. H.; Laun, J.; Junkers, T. Precise Polymer Synthesis by Autonomous Self-Optimizing Flow Reactors. *Angew. Chem., Int. Ed.* **2019**, *58* (10), 3183−3187.
(16) Cao, L.; Russo, D.; Felton, K.; Salley, D.; Sharma, A.; Keenan, G.; Mauer, W.; Gao, H.; Cronin, L.; Lapkin, A. A. Optimization of Formulations Using Robotic Experiments Driven by Machine Learning DoE. *Cell Reports Physical Science* **2021**, *2* (1), 100295.
(17) Vriza, A.; Chan, H.; Xu, J. Self-Driving Laboratory for Polymer Electronics. *Chem. Mater.* **2023**, *35* (8), 3046−3056.
(18) Gongora, A. E.; Xu, B.; Perry, W.; Okoye, C.; Riley, P.; Reyes, K. G.; Morgan, E. F.; Brown, K. A. A Bayesian experimental autonomous researcher for mechanical design. *Science Advances* **2020**, *6* (15), No. eaaz1708.
(19) Bateni, F.; Sadeghi, S.; Orouji, N.; Bennett, J. A.; Punati, V. S.; Stark, C.; Wang, J.; Rosko, M. C.; Chen, O.; Castellano, F. N.; Reyes, K. G.; Abolhasani, M. Smart Dope: A Self-Driving Fluidic Lab for Accelerated Development of Doped Perovskite Quantum Dots. *Adv. Energy Mater.* **2024**, *14* (1), 2302303.
(20) Lee, J.; Mulay, P.; Tamasi, M. J.; Yeow, J.; Stevens, M. M.; Gormley, A. J. A fully automated platform for photoinitiated RAFT polymerization. *Digital Discovery* **2023**, *2* (1), 219−233.
(21) Politi, M.; Baum, F.; Vaddi, K.; Antonio, E.; Vasquez, J.; Bishop, B. P.; Peek, N.; Holmberg, V. C.; Pozzo, L. D. A high-throughput workflow for the synthesis of CdSe nanocrystals using a sonochemical materials acceleration platform. *Digital Discovery* **2023**, *2* (4), 1042−1057.
(22) Beaucage, P. A.; Martin, T. B. The Autonomous Formulation Laboratory: An Open Liquid Handling Platform for Formulation Discovery Using X-ray and Neutron Scattering. *Chem. Mater.* **2023**, *35* (3), 846−852.
(23) Rahmanian, F.; Fuchs, S.; Zhang, B.; Fichtner, M.; Stein, H. S. Autonomous millimeter scale high throughput battery research system. *Digital Discovery* **2024**, *3* (5), 883−895.
(24) Kusne, A. G.; Yu, H.; Wu, C.; Zhang, H.; Hattrick-Simpers, J.; DeCost, B.; Sarker, S.; Oses, C.; Toher, C.; Curtarolo, S.; Davydov, A. V.; Agarwal, R.; Bendersky, L. A.; Li, M.; Mehta, A.; Takeuchi, I. On-the-fly closed-loop materials discovery via Bayesian active learning. *Nat. Commun.* **2020**, *11* (1), 5966.
(25) Ament, S.; Amsler, M.; Sutherland, D. R.; Chang, M.-C.; Guevarra, D.; Connolly, A. B.; Gregoire, J. M.; Thompson, M. O.; Gomes, C. P.; van Dover, R. B. Autonomous materials synthesis via hierarchical active learning of nonequilibrium phase diagrams. *Science Advances* **2021**, *7* (51), 12.
(26) Lo, S.; Baird, S. G.; Schrier, J.; Blaiszik, B.; Carson, N.; Foster, I.; Aguilar-Granda, A.; Kalinin, S. V.; Maruyama, B.; Politi, M.; Tran, H.; Sparks, T. D.; Aspuru-Guzik, A. Review of low-cost self-driving laboratories in chemistry and materials science: the "frugal twin" concept. *Digital Discovery* **2024**, *3* (5), 842−868.
(27) Martin, T. B.; Audus, D. J. Emerging Trends in Machine Learning: A Polymer Perspective. *ACS Polymers Au* **2023**, *3* (3), 239−258.
(28) Heil, C. M.; Patil, A.; Dhinojwala, A.; Jayaraman, A. Computational Reverse-Engineering Analysis for Scattering Experiments (CREASE) with Machine Learning Enhancement to Determine Structure of Nanoparticle Mixtures and Solutions. *ACS central science* **2022**, *8* (7), 996−1007.
(29) Treece, B. W.; Kienzle, P. A.; Hoogerheide, D. P.; Majkrzak, C. F.; Losche, M.; Heinrich, F. Optimization of reflectometry experiments using information theory. *J. Appl. Crystallogr.* **2019**, *52* (1), 47−59.
(30) Durant, J. H.; Wilkins, L.; Cooper, J. F. K. Optimizing experimental design in neutron reflectometry. *J. Appl. Crystallogr.* **2022**, *55* (4), 769−781.
(31) Kjær, E. T. S.; Anker, A. S.; Weng, M. N.; Billinge, S. J. L.; Selvan, R.; Jensen, K. M. Ø. DeepStruc: towards structure solution from pair distribution function data using deep generative models. *Digital Discovery* **2023**, *2* (1), 69−80.
(32) Walsh, D. J.; Zou, W.; Schneider, L.; Mello, R.; Deagen, M. E.; Mysona, J.; Lin, T.-S.; de Pablo, J. J.; Jensen, K. F.; Audus, D. J.; Olsen, B. D. Community Resource for Innovation in Polymer Technology (CRIPT): A Scalable Polymer Material Data Structure. *ACS Central Science* **2023**, *9* (3), 330−338.
(33) Schneider, L.; Walsh, D.; Olsen, B.; de Pablo, J. Generative BigSMILES: an extension for polymer informatics, computer simulations & ML/AI. *Digital Discovery* **2024**, *3* (1), 51−61.
(34) Lu, S.; Jayaraman, A. Pair-Variational Autoencoders for Linking and Cross-Reconstruction of Characterization Data from Complementary Structural Characterization Techniques. *JACS Au* **2023**, *3* (9), 2510−2521.
(35) Sutliff, B. P.; Goyal, S.; Martin, T. B.; Beaucage, P. A.; Audus, D. J.; Orski, S. V. Correlating Near-Infrared Spectra to Bulk Properties in Polyolefins. *Macromolecules* **2024**, *57* (5), 2329−2338.
(36) Edwards, C. E. R.; Lakkis, K. L.; Luo, Y.; Helgeson, M. E. Coacervate or precipitate? Formation of non-equilibrium microstructures in coacervate emulsions. *Soft Matter* **2023**, *19* (45), 8849−8862.
(37) Lee, S.; Bluemle, M. J.; Bates, F. S. Discovery of a Frank-Kasper σ Phase in Sphere-Forming Block Copolymer Melts. *Science* **2010**, *330* (6002), 349−353.
(38) Kim, K.; Arora, A.; Lewis, R. M.; Liu, M.; Li, W.; Shi, A.-C.; Dorfman, K. D.; Bates, F. S. Origins of low-symmetry phases in asymmetric diblock copolymer melts. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (5), 847−854.
(39) Biswas, A.; Liu, Y.; Creange, N.; Liu, Y.-C.; Jesse, S.; Yang, J.-C.; Kalinin, S. V.; Ziatdinov, M. A.; Vasudevan, R. K. A dynamic Bayesian optimized active recommender system for curiosity-driven partially Human-in-the-loop automated experiments. *npj Computational Materials* **2024**, *10* (1), 29.
(40) U.S. National Science Foundation: Award Abstract # 2152210 Graduate Traineeship on Advances in Materials Science using Machine Learning. https://www.nsf.gov/awardsearch/showAward?AWD_ID=2152210 (accessed 2024-08).
(41) Adorf, C. S.; Ramasubramani, V.; Anderson, J. A.; Glotzer, S. C. How to Professionally Develop Reusable Scientific Software—And When Not To. *Computing in Science & Engineering* **2019**, *21* (2), 66−79.
(42) Zhao, Z.; Chavez, T.; Holman, E. A.; Hao, G.; Green, A.; Krishnan, H.; McReynolds, D.; Pandolfi, R. J.; Roberts, E. J.; Zwart, P. H.; Yanxon, H.; Schwarz, N.; Sankaranarayanan, S.; Kalinin, S. V.;

https://doi.org/10.1021/acs.macromol.4c01410
Macromolecules XXXX, XXX, XXX−XXX

Mehta, A.; Campbell, S. I.; Hexemer, A. MLExchange: A web-based platform enabling exchangeable machine learning workflows for scientific studies. In *2022 4th Annual Workshop on Extreme-scale Experiment-in-the-Loop Computing (XLOOP)*; IEEE, 2022.
(43) *Hugging Face*. https://huggingface.co (accessed 2024-08).
(44) Pelkie, B. G.; Pozzo, L. D. The laboratory of Babel: highlighting community needs for integrated materials data management. *Digital Discovery* **2023**, *2* (3), 544−556.
(45) McDannald, A.; Frontzek, M.; Savici, A. T.; Doucet, M.; Rodriguez, E. E.; Meuse, K.; Opsahl-Ong, J.; Samarov, D.; Takeuchi, I.; Ratcliff, W.; Kusne, A. G. On-the-fly autonomous control of neutron diffraction via physics-informed Bayesian active learning. *Applied Physics Reviews* **2022**, *9* (2), 021408.
(46) Audus, D. J.; McDannald, A.; DeCost, B. Leveraging Theory for Enhanced Machine Learning. *ACS Macro Lett.* **2022**, *11* (9), 1117−1122.
(47) Ethier, J. G.; Audus, D. J.; Ryan, D. C.; Vaia, R. A. Integrating theory with machine learning for predicting polymer solution phase behavior. *Giant* **2023**, *15*, 100171.
(48) Raissi, M.; Perdikaris, P.; Karniadakis, G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **2019**, *378*, 686−707.
(49) Qu, C.; Schneider, B. I.; Kearsley, A. J.; Keyrouz, W.; Allison, T. C. Applying graph neural network models to molecular property prediction using high-quality experimental data. *Artificial Intelligence Chemistry* **2024**, *2* (1), 100050.
(50) *University of Toronto Acceleration Consortium*. https://acceleration.utoronto.ca/ (accessed 2024-06).
(51) *CASE: The Community for Autonomous Scientific Experimentation*. https://autonomous-discovery.lbl.gov/ (accessed 2024-06).
(52) Kiefer, K.; Rossa, L.; Wutzler, F.; Ektrom, N.; Pettersson, A.; Faulhaber, E.; Zolliker, M. An introduction to SECoP − the sample environment communication protocol. *Journal of Neutron Research* **2020**, *21*, 181−195.
(53) Wierenga, R. P.; Golas, S. M.; Ho, W.; Coley, C. W.; Esvelt, K. M. PyLabRobot: An open-source, hardware-agnostic interface for liquid-handling robots and accessories. *Device* **2023**, *1* (4), 100111.

https://doi.org/10.1021/acs.macromol.4c01410
Macromolecules XXXX, XXX, XXX−XXX