



Cite this: DOI: 10.1039/d5lp00380f

Convergence of high throughput experimentation and machine learning to rapidly advance application-specific polymer development

Mashrafee Aryan,^{†a} Daniel Struble,^{†a} Felix Campbell,^{id b} Saroj Upreti,^a
S. M. Ashik Abedin,^e Aahil Khambawla,^{id c} Jeetain Mittal,^{id c,d}
Michael S. Dimitriyev,^{id e} Emily B. Pentzer,^{id d,e} Svetlana A. Sukhishvili,^{id e}
Xiaodan Gu^{id a} and Boran Ma^{id *a}

Machine learning (ML) has heavily influenced the way scientific study is done with demonstrated successes in nearly every field. However, furthering performance and explainability of ML models in increasingly complex systems and with increasingly demanding outcomes requires a significant influx of high-quality data. To that end, this Review covers some of the techniques, instrumentation, and methodologies that have shown promise for significantly accelerating the discovery of polymer materials, optimization of their properties, and elucidation of property–application relationships through high throughput (HT) experimentation, characterization, and analysis. Attention is given to not only ML, but also to hardware advancements and their synergy with computational tools. Multiple studies are highlighted that demonstrate effective implementation of HT synthetic, data acquisition, or analytical methods, often with full integration of ML for the creation of fully autonomous workflows. We present our outlook and perspectives on the incorporation of HT techniques for the discovery and study of polymer materials within a broad range of applications, and include practical considerations for implementing HT methods at the laboratory scale.

Received 27th November 2025,
Accepted 26th May 2026

DOI: 10.1039/d5lp00380f

rsc.li/rscapppolym

1 Introduction

Polymer materials can be found everywhere from low-cost single-use items to high-performance electronics,^{1,2} bio-medical devices,^{3,4} coatings,⁵ thermal protection systems,^{6,7} and building materials.⁸ The ubiquity of these materials is due to their tunability that is driven by chemical, structural, and topological factors. The “Small Molecule Universe” is estimated to contain over 10⁶⁰ synthesizable compounds,⁹ and those are only compounds limited to 500 Da. The polymer materials design space is at least as large based solely on chemical composition. Beyond composition, structural features such as molecular weight and dispersity introduce additional emergent properties and applications.^{10–12} Furthermore, polymer materials include thermoplastics,

rubbers, gels, and thermosets with varying topological structures including branches, crosslinks, meso- and macrocycles, and bottlebrushes.^{13–16} While the range of design parameters for polymer materials makes them candidates across nearly every application, it also makes exploration of the entire design space and understanding the processing–structure–property relationships an outstanding challenge.

While substantial progress has been made toward the rational design of polymer materials and deconvoluting the relationships between structure, processing, and properties, a number of significant roadblocks still exist. Of those, the lack of standardized, accessible, reusable data has been the most critical barrier to data-driven discovery.¹⁰ However, recent advancements in high throughput (HT) methods for polymer synthesis, characterization, and property modeling and prediction, enabled by improved access to robotics and automation tools along with coding and machine learning (ML) algorithms have shown promise for accelerating the discovery and design of applied polymer materials.^{17,18}

ML encompasses a variety of techniques that can be broadly classified into classification, regression, and clustering tasks.¹⁹ This diversity of modeling approaches, combined with relatively simple implementation enabled by advances in modern computation, makes ML an extremely useful and ver-

^aSchool of Polymer Science and Engineering, University of Southern Mississippi, Hattiesburg, MS, 39406, USA. E-mail: boran.ma@usm.edu

^bDepartment of Chemistry, The University of the South, Sewanee, TN, 37383, USA

^cArtie McFerrin Department of Chemical Engineering, Texas A&M University, College Station, TX, 77843, USA

^dDepartment of Chemistry, Texas A&M University, College Station, TX, 77843, USA

^eDepartment of Materials Science and Engineering, Texas A&M University, College Station, TX, 77843, USA

[†]These authors contributed equally.



satellite tool for tackling complex, high-dimensional data that is common in polymer science and engineering. In the last 10 years alone, publications using ML for polymer science have increased more than 100-fold (research papers in 2025 *vs.* 2015, Web of Science, keywords: polymer, 'machine learning'), demonstrating the transformative nature of this tool for polymer scientists.

However, the Achilles' Heel of many ML models is their data-hungry nature, with many models requiring at minimum hundreds of data points to achieve reasonable performance. Relying solely on historic data or databases presents issues with accessibility, data provenance, and lacking or missing metadata. In addition, data of unknown quality or non-reproducible data may be unknowingly introduced.^{20–22} Reliable polymer data presents additional complexities due to factors such as stochastic structure, dispersity, processing history dependence, and complex synthetic pathways that are often inadequately documented in available data.^{11,12} To that end, developing HT workflows to synthesize, characterize, and model polymers and polymeric materials is crucial for advancing and accelerating study for applied polymers.

Traditional synthesis, characterization, and analysis methods are extremely slow, labor-intensive, and expensive by HT/ML standards.²³ Many recent studies have increasingly focused on the development of HT experimentation platforms for accelerated workflows, often paired with ML modeling and ML-guided automation. HT platforms can synthesize and characterize large libraries of polymers in a short period of time with minimal human input. With ML models, researchers can accelerate or automate characterization tasks and create predictive models for polymer properties. By using large datasets generated from HT workflows integrated with ML prediction framework, it is now possible to build increasingly autonomous, closed-loop platforms for polymeric materials discovery and optimization.^{24–28}

The goal of this Review is to highlight recent developments in both HT methods that are of particular importance for polymer science, as well as their incorporation with ML to accelerate the study and innovation of polymer materials across a range of application spaces. Our focus on the synergy of HT/ML and its impact within a wide range of practical applications compliments a number of other excellent reviews on other HT techniques and strategies,^{31–34} ML for polymer science,^{35,36} and on HT/ML for specific applications like polymer therapeutics,³⁷ biomaterials,³⁸ and organic solar cells.³⁹

The remainder of this Review is organized and discussed in four sections (Fig. 1) related to (1) HT techniques and instrumentation for polymer synthesis, (2) HT approaches to characterization (including data acquisition and analysis), (3) the synergy of HT techniques with ML, and (4) showcase studies of applied polymers. Individual sections build upon one another, as HT synthesis necessitates HT measurement, which necessitates HT analysis and interpretation. Together, these approaches can generate large volumes of data ideal for ML modeling.

Ultimately, combining HT experimentation with ML can enable closed-loop workflows and increasingly autonomous

laboratory systems to accelerate experimentation and discovery for applied polymers. It should be noted that, at this time, few systems exist without bottlenecks, and that this idealized workflow of HT synthesis, HT characterization, ML, repeat, poses many obstacles that are yet to be solved not only from a technical perspective, but also from financial, logistical, and human capital perspectives. Recently, fully closed-loop systems incorporating all HT and ML components have started to be realized,^{40,41} but broad application of fully HT/ML workflows remains aspirational for many labs and systems. This Review, therefore, focuses on many state-of-the-art techniques and instrumentation, often independent of a wholly HT or closed-loop system. At the end of this Review, we provide our perspectives and outlook for future directions of this field, including potential pathways to more widespread use and integration of the techniques and instrumentation highlighted herein.

2 High throughput synthetic techniques for polymers

HT methods are critical for rapid investigation and discovery of novel polymers and their materials. This section looks at some of the most common methods for accelerating the synthesis of polymers. It must be noted that "high throughput" is in many ways a moving target. To some labs, 96-well plate chemistry would be considered HT, while for others, multiple simultaneous 96-well plates are routine. Because of this variability, HT does not have a single, universal definition or quantitative threshold that can be applied to all scientific domains, or even all polymer science domains. Instead, the term is context-dependent and its interpretation can be shaped by domain definitions. Broadly, HT methods can be defined as the ability to generate and/or evaluate large numbers of samples in a short period of time. However, there is also no agreement on what defines a "large" number of samples or a "short" period of time. For instance, in biological HT screening, HT can mean evaluation of hundreds of thousands to millions of samples per day,⁴² whereas in scattering experiments, HT can mean the ability to screen tens to hundreds of samples per day.⁴³ For our purposes, HT is used as a descriptive term for methods that involve automation, parallelization, and/or combinatorial experimentation, typically facilitated or accelerated by computation.

2.1 Flow polymer synthesis

Flow synthesis combines reagents in a tubular flow reactor *via* pumps or syringes^{44,45} as shown in Fig. 2. Precise flow reaction control is enabled by tuning the molar ratio of reactants, flow rate, reactor length, and temperature. In addition to controllable reaction conditions, flow-based synthesis also provides rapid heat transfer, homogeneous mixing, and the potential to integrate multi-step reactions in-line.^{46,47} These factors lead to shorter reaction times, more reproducible results, and ease of scalability when compared to batch synthesis.^{48,49} The control and HT capabilities of flow reactions for polymers can be



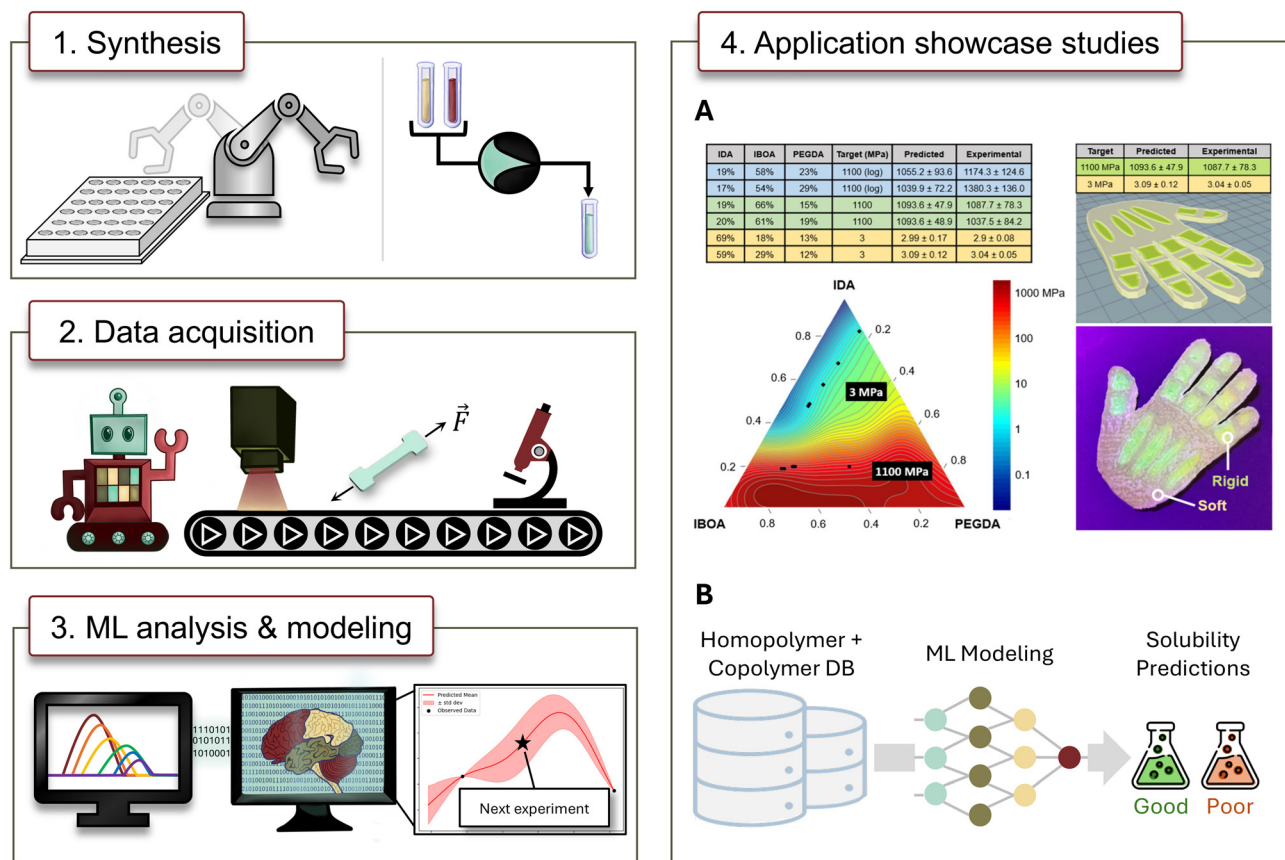


Fig. 1 Schematic overview of the experimental workflow: (1) synthesizing material, (2) taking measurements of properties of interest, (3) using ML to accelerate analysis, create models, and guide further experimentation, (4) use of new models or knowledge in application like (4A) rational design of materials with targeted modulus values or (4B) polymer solubility models. 4A adapted from ref. 29 with permission from the American Chemical Society, copyright 2024, licensed under CC-BY 4.0. 4B adapted from ref. 30 with permission from the Royal Society of Chemistry, copyright 2025, licensed under CC-BY-NC 3.0.

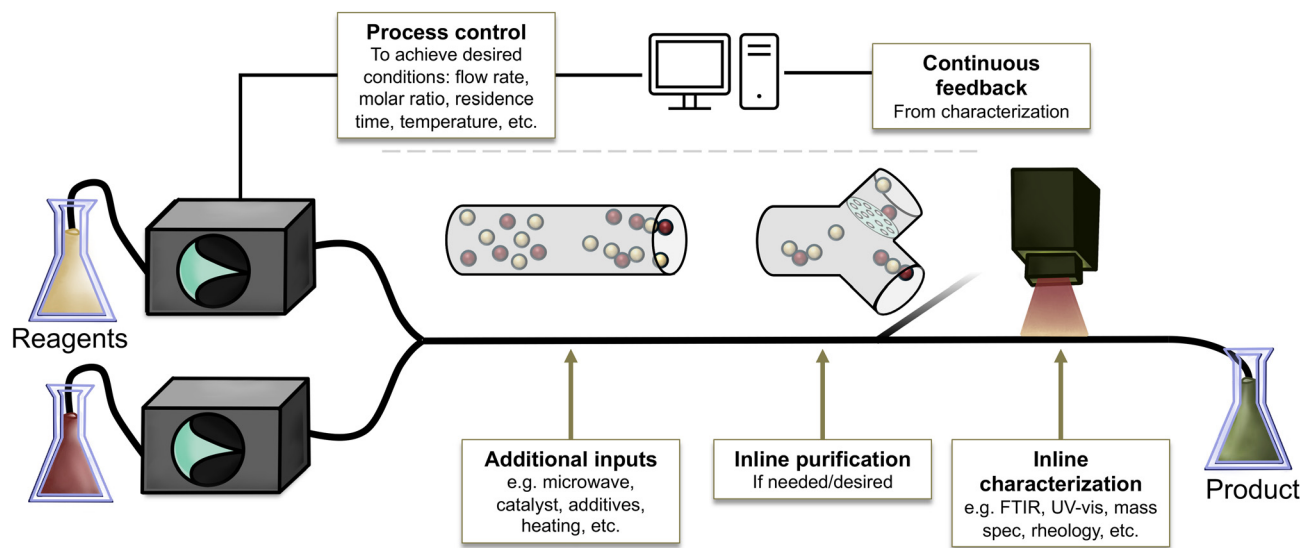


Fig. 2 Generalized schematic of a basic flow chemistry setup including elements that make the process especially high throughput and possibly automatable: in-line purification, in-line characterization, and continuous feedback that can control the process based on process data.



demonstrated by the industrial production of high-density polyethylene *via* loop-slurry reactors.⁵⁰ However, at the laboratory level, batch reactions are still more common than flow reactions. The reduced throughput and reproducibility in batch reaction can lead to bottlenecks for many projects where materials with excellent properties cannot be scaled up to the industrial level. The limitations of batch chemistry also hamper discovery of novel compounds by leaving large regions of experimental space underexplored or unexplored entirely. Converting from batch to flow can already improve throughput, but continuous flow synthesis can also follow a “numbering up” approach, that is, running multiple reactors in parallel and taking advantage of the ability to chain multi-step reactions efficiently. A shift toward laboratory use of flow reactors could provide a more efficient route for academia-to-industry transition. Flow synthesis for polymers has been recently demonstrated for various forms of reversible addition-fragmentation chain transfer (RAFT) polymerization,^{51–53} anionic polymerization,^{54–56} miniemulsion polymerization,^{57,58} and more.^{59–61}

One subfield that implements flow methods more commonly is the production of medical grade polymers. The consistency and fine-tuned control offered by flow synthesis is highly beneficial since batch-to-batch variability from traditional synthesis can lead to dramatically different outcomes for biological systems.⁶² This has been particularly relevant in recent years for drug delivery applications. Flow reactors have been used to fabricate polymers for drug delivery such as polymer-matrix nanomaterials,^{47,58,63} protein-loaded nanogels,⁶⁴ and core cross-linked polymeric micelles.⁶⁵ Studies that focused on optimizing consistency exhibited good control, demonstrating low particle polydispersity with reported values ranging from 0.1 to 0.001 (DLS-derived size distribution index),^{63,65} as well as controllable morphology through augmentation of flow conditions.⁵⁸ One study reported a 12× throughput increase when converting from batch to flow.⁶³ In another study, monomer conversion increased from 63% in batch to 97% in flow over an identical time frame due to faster flow reaction kinetics.⁴⁷ It was posited that these kinetics could be attributed to higher mixing rates and shorter path lengths present in flow reactors.⁶⁶ These studies on medical-grade polymers highlight key advantages of flow synthesis, including increased throughput and high purity, which could be extended to other areas of applied polymer research.

While flow polymerization offers significant advantages, there are limitations to be considered. Crosslinked, insoluble, or high molecular weight polymers can lead to clogging of flow reactors because of high viscosity or the formation of precipitates. Chanchaona *et al.* reported the flow synthesis of hypercrosslinked polymers and difficulties faced due to the formation of insoluble products and solvent adsorption. However, by optimizing reactor design and conditions, these limitations were overcome and they reported 32- to 117-fold higher productivity than comparable batch reactions.⁶⁷ Additionally, reactor materials could face incompatibility with reactants, solvents, or catalysts.⁴⁷ One study demonstrated the

fabrication of a reactor with perfluoropolyether and subsequent thermal treatment, providing high chemical resistance that allows for the highly reactive RAFT polymerization.⁴⁷ Another concern is that scaling up individual reactors can reduce the consistency of flow-based systems. For example, increasing reaction chamber diameters lowers heat transfer and can lead to different or varying mixing conditions that affect the formation of desired product.⁶⁸ This, again, is why numbering-up is a more common route to scale up flow systems.

2.2 Robotics and automation

Robotic equipment is an alternative, and sometimes overlapping, option for accelerating polymer synthesis that is increasingly being integrated into academic laboratories. Automation of laboratories reduces the burden of manual and repetitive tasks on researchers while consistently demonstrating higher throughput and improved precision. It also enables the chaining of multiple processes together, creating efficient workflows. In addition, it can reduce human contact with potentially harmful processes, reactants, and intermediates. For these reasons, the presence of automation continues to grow in laboratories.

Among the most common forms of automation are modular platforms, robotic arms, and liquid-handling robots. “Modular platform” is a term that can be applied to almost any form of in-lab automation, but more specifically it can be considered as a platform with tailorable structure and a variety of modules to choose from for customizable experimentation.⁶⁹ Modular platforms provide a basis for building up integrated synthesis, purification, processing, and/or characterization systems. These are found in polymer science for the production of polymer nanotubes,⁷⁰ drug delivery polymers,⁷¹ and a variety of other applications.^{72,73} Modular platforms for parallel synthesis can be based on compatibility with 96-well plates commonly used in the biosciences. 96-well plates have seen frequent use in polymer science due to their low cost, easy maintenance or replacement, and simple implementation.^{10,44,74}

Robotic arms provide the classical imagery of robotics integration into laboratories and have seen use in the development of application-driven polymers, such as integration into an automated polymer press, and the Polybot for electrical thin film processing.^{75–77} Robotic arms shine at repetitive, multi-step tasks where their precision and dexterity allows them to accomplish tasks comparable to human input.⁶⁹

Liquid-handling robots have been most widely used in life sciences.⁷⁸ They are particularly useful for parallel polymerization to generate polymer libraries^{37,79,80} for the exploration of huge chemical spaces to elucidate structure–property relationships (Fig. 3A). Polymer libraries have been made across a wide range of application spaces, including medicine,^{81,82} electronics,²⁴ and biodegradable plastics.⁸³ For the pharmaceutical industry, library creation is considered a critical step for the discovery of new candidate molecules.^{84,85}

There are many demonstrations of automated polymer library synthesis producing diverse polymers for characterization and use. Continuous flow with autonomous control of



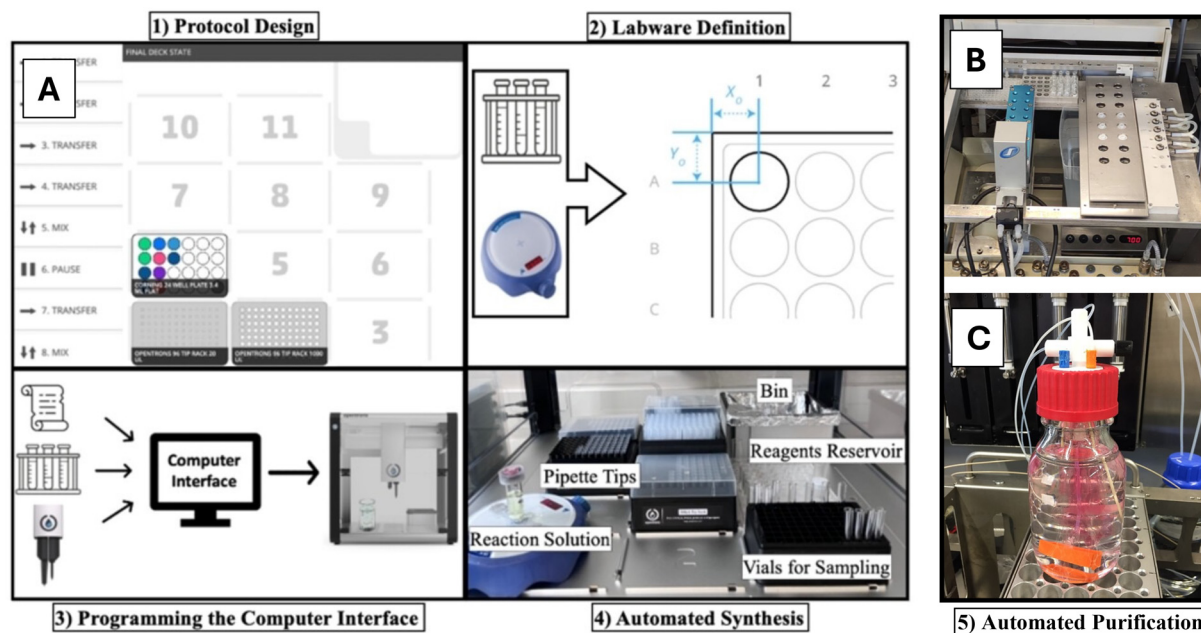


Fig. 3 Automation of key steps in the synthetic workflow. (A) “Four steps” of RAFT polymerization library creation *via* Openotrons liquid handling robot. Reproduced from ref. 79 with permission from Wiley-VCH GmbH, copyright 2023, licensed under CC BY-NC-ND 4.0. (B) Automated dialysis system attached to liquid handling robot. Adapted from ref. 86 with permission from MDPI, copyright 2022, licensed under CC-BY 4.0. (C) Automated continuous dialysis system integrated with a synthesis robot. Reproduced from ref. 87 with permission from MDPI, copyright 2020, licensed under CC-BY 4.0.

parameters has been demonstrated for the synthesis of a library of a hundred distinct block copolymers in just nine minutes.⁸⁸ Liquid-handling robots have been demonstrated for polymer library creation using various mechanisms such as RAFT polymerization and solid-phase synthesis, including for applications such as cholesterol-lowering drugs.^{37,79,80,89}

Common difficulties for robotic laboratory integration manifest in high upfront cost and limitations of current robotic hardware.¹⁰ Robotic integration for polymers specifically faces several unique challenges. Due to the high viscosity of polymers, clogging of micropipettes in liquid-handling robots and other similar systems can occur.⁹⁰ In addition, many polymer starting materials are solids (monomers, catalysts, fillers) or require processing (powder mixing, molding, pressing) for which commercially available robotics systems are unsuited.⁹¹ Certain types of polymerization are oxygen-sensitive, such as living/controlled polymerization, which cannot be used in conventionally open-air robotics systems.⁷⁹ Adapting off-the-shelf robotics platforms to the needs of a functional polymer laboratory often requires further costs, expertise, and customized equipment.

2.3 High throughput purification

The purifying of polymers is dynamic, as factors such as variable monomer conversion, side reactions, and variable amounts of byproducts necessitate different purification processes. There has been little work on the automation of polymer purification in HT workflows, creating a bottleneck for many automated processes.²⁵ The work that has been done

largely focuses on the ability to integrate with existing HT systems.

The most explored method of automated purification is dialysis. Dialysis can be automated with liquid-handling robotics⁸⁶ as well as flow systems^{92,93} (Fig. 3B and C). Flow systems may prove particularly attractive for pairing with dialysis, as higher dialysate flow rate has been demonstrated to improve clearance rate in other systems.⁹⁴ In addition to faster purification, dialysis has also been demonstrated to use less solvent than comparable manual approaches.^{77,87} Ultrafiltration is a flow-compatible, underexplored purification method that handles incomplete conversion while maintaining a reasonable throughput.⁹⁵ Additionally, gel permeation chromatography has been paired with a 96-well plate with 95% small molecule impurity removal reported.⁹⁶

In general, HT purification of polymers is an extremely underdeveloped field. In their user guide to high throughput workflows, Day *et al.* recommend automated synthesis of only high-conversion synthetic reactions to bypass the need for purification, which is commonly practiced with researchers often opting for slower reactions with complete conversion rather than those requiring the setup and optimization of an additional purification step.^{97,98} However, some approaches like those involved in biopolymer and sequence-defined polymer synthesis can be significantly simplified due to presence or absence of known functionalities that enable selective precipitation of the target polymers⁹⁹ or efficient removal of impurities and low-molecular-weight oligomers,¹⁰⁰ or by enabling simple fractionation by chromatography when target molecular weights are discrete.^{42,101}



3 Advancing instrumentation toward HT characterization

As HT synthesis continues to grow, equally significant advances in HT characterization have emerged. It should be noted that “high throughput” may not be directly quantifiable, but for our purposes refers to methods that involve automation, parallelization, and/or combinatorial experimentation, typically facilitated or accelerated by computation. Given the broad range of different characterization equipment and techniques used by polymer scientists, in-depth discussion of advancing HT methods across them all is not practical, but this section endeavors to highlight advances in HT methods within several of the most common.

3.1 Spectroscopy

Chemical composition analysis helps researchers elucidate chemical structures and identify functional groups. Fourier Transform Infrared (FTIR) spectrometers have been equipped with autosamplers, allowing them to rapidly scan large sample spaces to get a functional group mapping of polymer libraries.¹⁰² Study of microplastics in the environment has been a significant driver for HT spectroscopic techniques. Accelerated near-infrared (NIR) and Raman spectroscopy have both been adapted for HT while demonstrating success in the difficult task of identifying microplastics with complex heterogeneous samples.^{103,104} Nuclear Magnetic Resonance (NMR) spectrometers are commonly equipped with autosamplers, but they can also be adapted to flow-mode setups. Both options allow multi-sample NMR and flow NMR allows for in-line or on-line measurements as well.¹⁰⁵

3.2 Chromatography

Gel Permeation Chromatography (GPC), the most widely used method for polymer molar mass characterization, provides apparent molar mass distributions relative to calibration standards. Modern GPC systems are commonly equipped with autosamplers to improve throughput *via* automation. Microfluidic GPC platforms have also been developed, enabling faster analyses with reduced sample volumes. While the resolution is more limited compared to conventional size exclusion methods, these systems offer promising throughput advantages for screening applications.¹⁰⁶

Advancing traditional GPC, Murphy *et al.* have recently demonstrated automated systems capable of resolving dozens of narrowly dispersed polymer fractions that also include variations in block composition and molar mass, from a single parent polymer. This development has enabled the separation of large, well-defined libraries for downstream characterization.^{107–109}

3.3 Microscopy

Unlike traditional microscopy that focuses on a small sample set, HT microscopy techniques are designed to rapidly characterize multiple samples. Through integration of advanced imaging methods with automation, ML, and combinatorial

sample preparation, HT microscopy has enabled the exploration of polymer microstructure, morphology, and chemical compositions across a wide range of material libraries. This scalability has accelerated the understanding of structure–property relationships that ultimately guide rational materials design.^{40,60,110}

Depending on modality, microscopy can be broadly categorized into Optical Microscopy (OM), Confocal Fluorescence Microscopy (CFM), Electron Microscopy (EM), Atomic Force Microscopy (AFM), Super Resolution Microscopy (SRM), and others. CFM when integrated with HT methodology has found great use in analyzing the compositional heterogeneity of hundreds of olefin polymerization catalyst particles in both 2D and 3D.¹¹¹ Similarly, HT OM has enabled rapid mapping of phase behavior in polymer blends and coacervates, which is otherwise quite time-consuming with conventional approaches.^{112,113} High-resolution EM with recent advances like 4D-STEM (scanning transmission electron microscopy) has allowed for precise mapping of crystalline domains in polymers, directly revealing chain arrangements and lattice deformations at submolecular resolution.¹¹⁴

Recently, AFM has seen increased imaging speed and throughput without major compromise in spatial resolution. This has been made possible due to the development of high-speed dynamic mode AFM, adaptive multiloop-mode imaging, and bimodal AFM.^{115–117} The automation of AFM workflows through robotic sample handling, automated tip exchange, batch-mode imaging, ML, and advanced data analytics is increasingly being used for combinatorial, complex, heterogeneous polymer systems (Fig. 4).^{60,118,119} Further, the development of polymer-based and 3D-printed AFM tips and cantilevers accompanied by multipurpose 3DTIPs and self-actuated cantilevers has helped reduce tip wear and improve imaging in air or water environments.^{120,121}

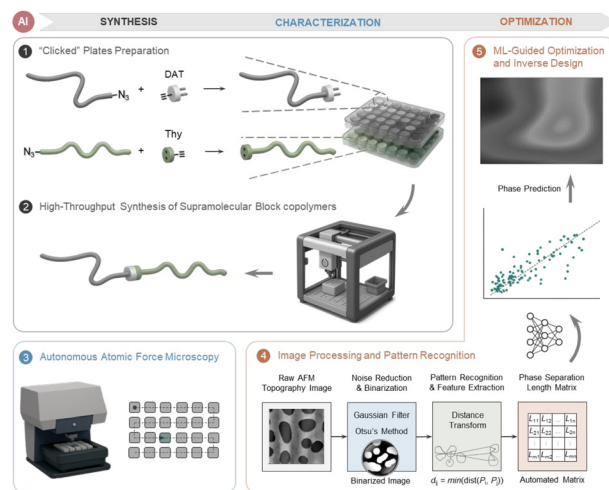


Fig. 4 Combinatorial, automated supramolecular polymer blends facilitated by robotic synthesis, automated AFM, and ML analysis and modeling techniques. Adapted from ref. 60 with permission from ChemRxiv, copyright 2025, licensed under CC-BY-NC-ND 4.0.



3.4 Light scattering

HT scattering techniques have sped up applied polymer research by enabling rapid and detailed characterization of polymer structure, dynamics, and properties across large sample sets and parameter spaces. These methods primarily include light (LS), X-ray (XRS), and neutron scattering (NS), which are being routinely integrated with automation,¹²² advanced data analysis,¹²³ and ML. This has consequently accelerated materials discovery, optimized processing, and deepened fundamental understanding of polymers.

Most of the lab-based XRS experiments are limited by inherently low energy X-ray sources, and typically require tens of minutes to hours for measurement of an individual sample.¹²⁴ Fortunately, modern synchrotron sources are able to deliver extremely intense, focused, and monochromatic X-ray beams allowing for collection of high-quality data in seconds to minutes.¹²⁵ Using these sources, small/wide-angle X-ray scattering (SAXS/WAXS) experiments have been carried out over shorter time frames to investigate the structural evolution of different polymer sample types such as conjugated polymers, block copolymers, *etc.* HT scattering data can then be leveraged to establish structure–property relationships that can lead to tailored polymer systems with targeted specific properties.^{40,110}

3.5 Mechanical characterization

Mechanical properties are critical for polymer applications. Mechanical properties such as tensile strength, hardness, and modulus are commonly studied and have seen advancements toward HT measurement systems. In mechanical testing, data acquisition can be intrinsically accelerated with the use of smaller samples that can reduce instrument translation times and the amount of strain required during deformation. For example, automated nanoindentation systems¹²⁶ can rapidly characterize hardness and elastic modulus with high-speed scanning to create modulus maps that show how properties vary within and across samples. This can be beneficial for characterizing libraries of new materials or formulations,¹²⁶ or to characterize dispersion in nanocomposites.¹²⁷ Small samples also require less material and can facilitate bulk fabrication of test specimens to accelerate mechanical testing workflows. Microtensile and miniaturized dynamic mechanical analysis systems test small samples to measure ductility and yield or rheological behavior. These systems allow HT testing relative to traditional tensile instruments.^{128–130} It bears noting that for polymers, viscoelastic properties such as creep and stress relaxation have long had a pseudo-HT analogue available in the form of time–temperature superposition. Master curves map short-term data to long-term behavior, making otherwise intractable regimes experimentally accessible.¹³⁰

Polymer rheology has also seen advancements toward HT characterization, with researchers employing a range of methods including electromechanical,¹³¹ centrifugal¹³² and even optical microscopy,^{112,133} typically accompanied with

computational analysis and often in conjunction with ML.¹³⁴ These techniques all use relatively small amounts of material, a benefit for complex or difficult-to-synthesize polymers. Some of these HT rheological techniques can even obtain high-resolution time-dependent data *in situ*, allowing alternative HT options for evaluating related properties or behavior like polymerization kinetics.¹³³

4 ML-guided HT data analysis and property prediction

HT methods for synthesis and data acquisition necessitate high-rate data analysis, modeling, and process control. ML excels in all these tasks and has emerged as a natural complement to HT experimental methods. This section will briefly address how ML assists in data analysis bottlenecks from HT experimentation, how ML can accelerate the overall rate of scientific discovery for polymers through sophisticated modeling and property prediction, and how ML and automation can be deployed synergistically toward fully automated self-driving laboratories.

4.1 ML overview

ML broadly refers to a wide range of statistical methods for regression, classification, and clustering tasks that are all common in scientific research. Regression is the statistical method for defining relationships between independent and dependent variables. For example, the Flory–Fox equation is a traditional model that describes glass transition temperature as a function of molecular weight.¹³⁵ While effective given its simplicity, the Flory–Fox equation breaks down for some polymer topologies, polymer blends, systems with significant hydrogen bonding or crosslinking, and other common experimental situations. ML, by contrast, can model complex, non-parametric relationships with as many descriptors as can be provided, including functional groups, topology, blend composition, and more to capture much more nuanced relationships between input variables and predicted properties.

While regression predicts continuous outputs based on inputs, classification refers to predicting discrete groups based on inputs. An example of classification is a model that predicts “soluble” or “insoluble” as binary classes. Classification models can also be non-parametric and high-dimensional and are able to account for chemical, topological, and environmental descriptors to classify the behavior of even complex solutions.

Clustering is the third common ML task in polymer science and is useful for identifying common groupings in large datasets. The most typical form of clustering is dimensionality reduction – creating 2D or 3D groupings of high-dimensional data using some form of grouping to maintain global spacing between points or to maintain local cluster relationships. Further discussion and introduction to ML can be found in literature.^{10,136–140}



As suggested herein, the most critical part of any ML model is the quality of its training data. ML models can be trained on experimental data, historic data curated from datasets, or a combination thereof. Database-derived datasets are typically larger than experimental datasets generated for a specific study, and often cover a larger range of inputs (*e.g.*, more chemical diversity), but they frequently lack essential experimental details (*e.g.*, reaction time, processing conditions, environmental humidity, *etc.*) or provenance required to evaluate data quality and reproducibility.

The process of training ML models varies from task to task but can generally be broken into the following steps (Fig. 5): (1) curate a dataset either experimentally or from database (s).¹⁴¹ (2) Preprocess data.^{142,143} This may include cleaning, scaling, or transforming, and determining appropriate tools for fingerprinting if needed, *e.g.*, RDkit,¹⁴⁴ Morgan Fingerprint,¹⁴⁵ *etc.* (3) Data splitting.¹⁴⁶ If a test set is being generated from the available data, known as a holdout set, that data must be sequestered from other training data. It may be necessary to stratify the test set selection or otherwise ensure it is representative of the full dataset,¹⁴⁷ particularly if the dataset is small. 5% and 10% of the total dataset are common holdout sizes. (4) Model training. Fitting the model to the training data is typically done using multiple instances of the same model architecture. This is usually accomplished by *k*-fold cross-validation, where *k* = 5 and *k* = 10 are common choices.¹⁴⁸ Cross-validation results in a more robust estimate of the model's performance and generalizability to unseen data, making it a more reliable choice for hyperparameter selection than simply training the model one time on all available data. Hyperparameter optimization is also done during this time, where different options for model architectures and learning behaviors can be specified to see what best captures the underlying patterns in the data. Hyperparameter optimization is typically conducted using random search, grid search, or an optimization approach like Bayesian optimization.¹⁴⁹ (5) Model evaluation. During training, the "best" model is evaluated using an appropriate loss function. For regression, the most

common loss functions are mean squared error and mean absolute error. For classification, cross-entropy loss and hinge loss are the most common loss functions. Once trained, overall model performance is typically evaluated using coefficient of determination (R^2), root mean squared error, and/or mean absolute error for regression; and accuracy, recall, and/or F1 score for classification.^{150,151} While cross-validation techniques should improve generalization of models, it is critical to bear in mind that models may overfit, particularly problematic when predicting on data not derived from the original dataset.

This overview covered some basic concepts for understanding the rest of the section, but ML includes a range of techniques and methodologies that are beyond the scope of this Review and often with specific considerations on a per-dataset or per-task basis. Readers interested in polymer-specific ML best practices are directed to other resources related to ML models,^{19,152} ML methods and tutorials,^{139,152} polymer representations and selections for ML,^{153–156} model validation strategies,^{152,156} uncertainty and extrapolation limits of ML models,^{157,158} explainable ML,^{159–161} and more detailed discussions on the state of the ML field with respect to polymers, specifically.^{10,36,162–164}

4.2 ML-assisted HT data analysis

4.2.1 Spectroscopy.

Spectroscopic methods such as FTIR, Raman spectroscopy, and UV-Vis are commonly used for the analysis of macromolecular composition. ML models can be trained to predict spectral features by correlating spectra with molecular structure, composition, or functional properties. Given the ubiquity of these techniques across small molecule, macromolecular, biological, and other sciences, fully generalizable ML models for spectral elucidation have not yet been demonstrated.¹⁶⁵ Rather, models can be trained using relatively smaller datasets for interpretation of spectra toward specific applications. For example, ML has been used to characterize biodiesel candidates¹⁶⁶ and for identifying the composition of environmental microplastics by deconvoluting FTIR^{167,168} or Raman¹⁰⁴ spectra. Recently, ML has also demonstrated the ability to determine average polymer molecular weight from FTIR spectra.¹⁶⁹

Given the complexity of fully elucidating characterization spectra, robust ML for comprehensive interpretation remains a challenge. However, in specific applications with limited need for generalization there has been some progress. For example, ML has identified sterile or contaminated cells,¹⁷⁰ and has been used to predict particular features within UV-Vis spectra that correlate with phototoxicity.¹⁷¹ A somewhat easier task, predicting spectra from compounds has been successfully demonstrated and shows promise as a screening tool. ML methods have been used to predict UV-Vis spectra for drug discovery¹⁷² and photodetector design.¹⁷³

4.2.2 Microscopy.

HT microscopy is capable of producing hundreds of images in hours to days, making manual interpretation and analysis of all the data unfeasible. Convolutional Neural Networks (CNNs) have dominated the image analysis-by-ML paradigm for the last few decades. The

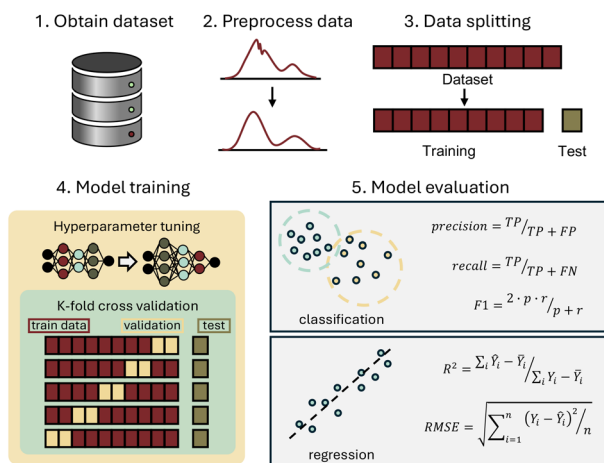


Fig. 5 Simplified overview of ML model training steps.



convolutional layers of a CNN utilize imaging “filters” to detect specific features from images like horizontal lines, vertical lines, textures, or curves.^{110,174} By stacking convolutional layers, CNNs can become extremely effective at extracting information from images, and they excel at classification and sometimes regression tasks based on image inputs (Fig. 6).¹⁷⁵ CNNs are a form of supervised learning – they rely on labeled data to determine what information they should be learning from images. CNNs have demonstrated the ability to evaluate miscibility of polymer blends from SEM images,¹⁷⁵ detect and estimate the size of polymer nanostructures in TEM,¹⁷⁶ and to classify block copolymer morphology from AFM images.¹¹⁰

4.2.3 Scattering. X-ray scattering data analysis can also be significantly accelerated by ML. Models have been trained to allow for rapid analysis and phase identification from scattering data^{179–181} and the development of *in situ* and *operando* scattering platforms for real-time monitoring of polymer processing and device operation.^{182–184} Consequently, complex, multiscale, and dynamic phenomena in polymers, such as phase separation, crystallization, and self-assembly could be studied effectively with unprecedented speed and resolution.^{185–187} However, challenges remain in data interpretation, model development, and the integration of scattering with other HT and computational methods.^{188–190}

Additionally, clever approaches to bypassing the difficulty of interpreting complex scattering data have been developed that leverage ML. The Jayaraman group has developed a Computational Reverse-Engineering Analysis for Scattering Experiments (CREASE) method that uses ML to both automate and accelerate the interpretation of complex scattering patterns (Fig. 7).^{191–194} ML models can also identify block copolymer phases from noisy data, including gyroid and σ phases, from scattering data without requiring any chemical information and in real time.¹⁸⁰

4.3 Property prediction

In addition to HT data analysis, ML is also commonly employed for modeling and subsequently predicting polymer properties based on chemical, structural, and/or processing inputs. For complex, interdependent, high-dimensional systems ML has emerged as a powerful modeling tool that is well-suited for large, multivariate datasets and flexible, non-parametric relationships – that is, models that learn relationships directly from data rather than assuming predefined forms such as linear or polynomial as with traditional regression methods. From an experimental design standpoint, ML does not necessarily require careful planning of experimental combinations like Design of Experiments (DOE) does, making it more flexible to implement in real-world conditions where orthogonality, randomization, and pre-existing knowledge of design space boundaries may be difficult to achieve. That said, DOE principles can still prove valuable for iterative modeling and selection of data points with which to train ML models.^{195–197} DOE and ML have been implemented in complementary or synergistic ways,^{198–201} though to our knowledge no specific instance of such a study or workflow exists related to the modeling or development of applied polymers. As stated, ML has the additional advantage over DOE in that it can access non-parametric models, improving flexibility and the ability to capture complex relationships.

ML algorithms have been used to predict polymer-related properties that have eluded more traditional modeling techniques, such as polymer solubility,^{30,202–204} glass-transition temperatures,^{205,206} dielectric constants,^{207–209} and more. Several of these will be highlighted in the showcase studies section of this Review. However, the broad range of available model architectures necessitates careful decision-making and awareness of the underlying data. For example, T_g prediction is

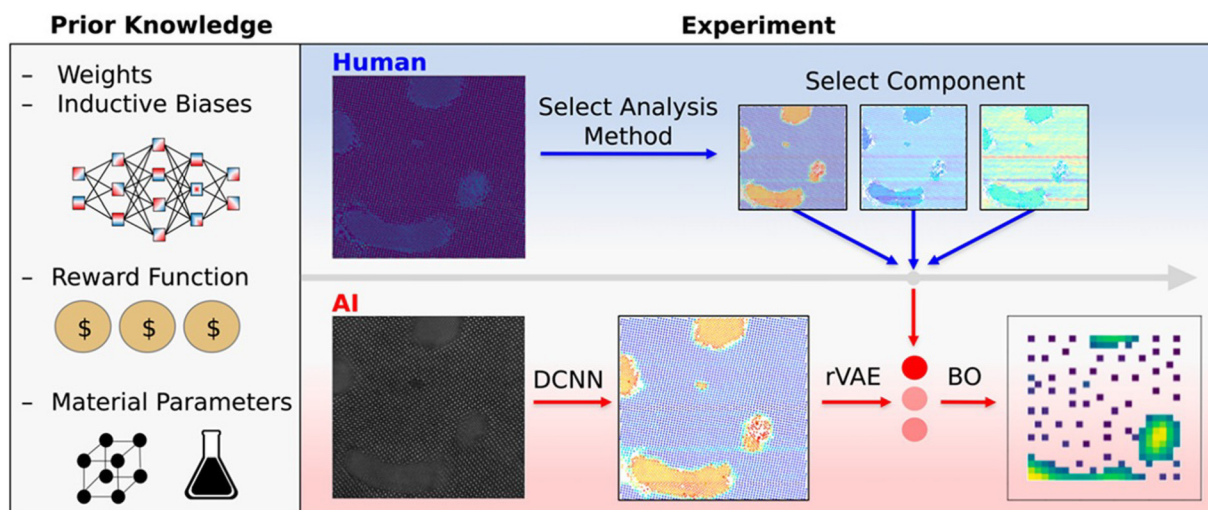


Fig. 6 Example workflow for an automated microscopy pipeline with in-line ML. Reprinted with permission from ref. 177 with permission from the American Chemical Society, copyright 2021. Original elements (STEM imaging) adapted from ref. 178 with permission from Springer Nature, copyright 2021, licensed under CC-BY 4.0.



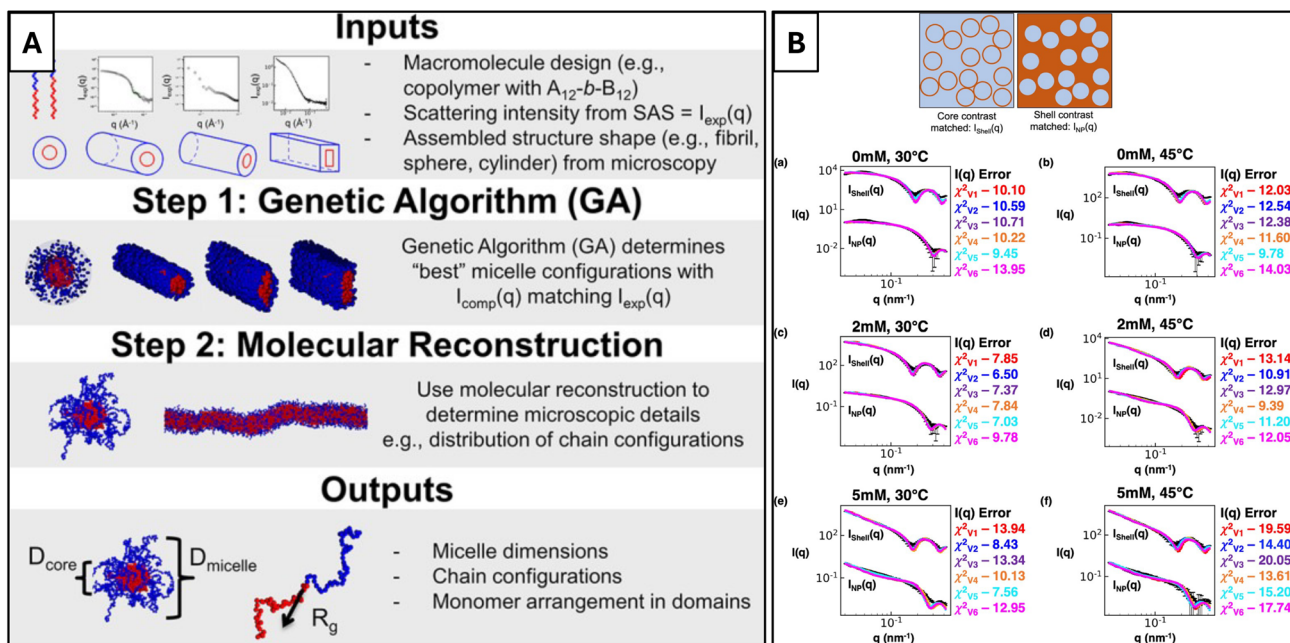


Fig. 7 CREASE scattering elucidation. (A) Overview of the CREASE workflow. Reproduced from ref. 179 with permission from the American Chemical Society, copyright 2021. (B) CREASE applied to small-angle neutron scattering profiles where black curves are experimental and colored curves are from CREASE. Reproduced from ref. 191 with permission from the American Chemical Society, copyright 2023. Both figures licensed under CC-BY-NC-ND 4.0.

a common task for polymer ML models. Researchers have had success on this task using all sorts of models and datasets,^{205,206,210} but a comprehensive benchmark study by Tao *et al.* demonstrated that for a given task and dataset, not all ML architectures and modeling decisions are made equal.¹⁵³ As ML modeling becomes increasingly integrated into polymer science, we emphasize that making ML models is easy, but making good ML models is challenging.

4.4 Closed-loop workflows and self-driving laboratories

Closed-loop workflows demonstrate the current pinnacle of combining HT and ML techniques into unified methodology, and emerge when ML models form a feedback loop with data acquisition methods. Summarizing the Review to this point: HT synthetic and purification techniques naturally require HT characterization and analysis. Then, giving that information to ML for system modeling and decision-making comprises a full HT/ML workflow. When ML is then leveraged to guide decision-making or further experimental design, the system enters a feedback loop known as a closed-loop workflow. The following sections use terms like "autonomous" and "self-driving". These terms are roughly synonymous in the literature, but in our discussion we use several terms as follows: "automated" meaning part or parts of the process are performed automatically by robotics or computer-driven systems. This can include autosamplers, automated GPC, or other systems that can perform a repetitive, predetermined task. "Autonomous" refers to systems that perform an automated task in response to some input. For example, an instrument

with in-line monitoring that can adjust its settings to maintain a target output or property. "Self-driving" refers to systems that have automated and autonomous elements, but that also engage in decision-making, that is, autonomous and closed-loop systems.

4.4.1 Closed-loop workflows. Beyond its uses for analyzing data for regression, classification, and clustering, ML can also be used as a powerful decision-making tool. In particular, active learning (AL) methods can be used to suggest subsequent experiments driven by desired outcomes by creating models based on known data, then determining where additional data is needed. Because these suggestions bridge the gap between testing an iteration and intelligently designing the next steps, they serve to "close the loop" of the design-build-test-learn (DBTL) cycle by strategically suggesting the next design. AL models employ three primary strategies: exploration, exploitation, or a balance of the two. Exploration refers to selecting data for acquisition that reduces global model uncertainty, while exploitation prioritizes sampling in regions believed to be near the optimum for the given task. Multiple ML strategies for AL exist, but one of the most common is using Bayesian Optimization (BO)²¹¹ with Gaussian Process Regression (GPR).²¹² GPR in this case is the actual ML model, which is a common choice for BO since it intrinsically provides uncertainty measurements for its predictions. Those uncertainty measurements can then be used for BO, which is not an ML model, but rather a sequential design strategy useful for selecting subsequent high-value data points. Using these approaches, AL can navigate regions of experimental space that go beyond



simple chemical intuition or heuristics and find novel solutions to optimization challenges. This ability to intelligently and efficiently explore the design space using ML guidance leads to the natural convergence of HT synthesis techniques, data acquisition methods, and ML-enabled analysis and modeling in closed-loop workflows.

4.4.2 Self-driving laboratories. Full integration of AL with automation gives rise to self-driving laboratories (SDLs) – automated systems that engage in the DBTL cycle with minimal or no human interaction.²¹³ SDLs differ from traditional automation as previously described, in that they are dynamic systems that alter their workflow in real time, allowing autonomous design, execution, and analysis of experiments.^{25–28}

By allowing AL models to suggest experiments that will further improve model performance, human intervention can be minimized in the DBTL loop. Letting integrated systems run autonomously can improve throughput⁵⁹ and consistency. The addition of AL methods can also improve data efficiency by selecting statistically optimized experiments. The ability of AL to intelligently design and implement experiments is the final component to enabling truly closed-loop automated workflows that are suggested by HT experimental systems. Several recent works using SDLs for applied polymers have focused on multi-objective optimization of several polymer nanoparticle properties such as maximum monomer conversion and minimized dispersity,²¹⁴ optimization of electronic thin film processing conditions (Fig. 8),^{24,28} and design of copolymers for enzyme stabilization.²¹⁵ While AL methods can be built from the ground up, workflows have also been democratized to make using these tools more accessible than ever as in the case with the gpCAM library that is both user-friendly for new practitioners of AL and robust to more demanding needs for advanced users.²¹⁶

5 Showcase studies

This section will highlight several studies across a range of application spaces for polymers. Each of these studies demonstrates multiple or all of the HT methodologies covered in this Review: HT synthesis, purification, and/or data acquisition paired with ML methods for analysis and/or modeling and prediction. Many of these showcase studies also use ML modeling and HT techniques in concert to create closed-loop or autonomous workflows.

5.1 Polymers for drug delivery

Polymers provide unique design opportunities due to their size and tunability, which can impart stimuli-responsive or self-assembly behaviors. Leveraging these behaviors, polymers are attractive options for drug delivery applications. In high-sensitivity applications like drug delivery, automated control and improved throughput facilitated by ML has demonstrated great progress.

Recent work by Xu *et al.* presented an SDL that optimizes the lower critical solution temperature (LCST) of poly(*N*-iso-

propylacrylamide) (PNIPAM).²¹⁷ PNIPAM exhibits a sharp and reversible phase transition near physiological temperature which makes it an ideal candidate for stimuli-responsive biomaterials applicable as drug delivery devices, artificial tissue scaffolds, and biosensors. The authors built a low-cost, modular platform that integrates robotic liquid-handling for precise formulation with BO that uses a GPR surrogate model. The system prepared many solution compositions in parallel, then measured cloud point transitions with automated optics and fed the results to the optimizer. Within only a few cycles, the platform converged on user-defined LCST targets in both two-salt and three-salt solution systems while making efficient use of experiments. The study showed how closed-loop HT/ML experimentation can be used to control a thermal transition with direct bearing on biomedical function. It also provides an accessible blueprint for autonomous experimentation that can be adapted to other thermoresponsive polymers where a specific transition window is required for device performance.

The Soft Materials Research and Technology (SMART) Lab at the University of Pennsylvania has used HT/ML to fabricate microfluidic double emulsion droplets. These polymer droplets are biomimetic with encapsulating layers allowing for protection of inner materials while external structure allows for integration into cells. These materials are commonly recognized to have potential for drug delivery applications; however, these cells require precision manufacturing, subject to a wide design space and highly sensitive fabrication and processing. In medical grade materials, even slight imperfections can lead to fatal consequences. O'Callaghan *et al.* utilized microscale flow chemistry to drastically limit batch-to-batch variability of polymer-based protocells, reporting the consistent formation of microcapsules as small as 14 microns – the smallest to date at that time.²¹⁸ More recently, the SMART laboratory has utilized the ML-empowered Automated Double Emulsion Droplet Library generator (ADLib) to evaluate microdroplets in real-time with ML.²¹⁹ They used ML in the form of automated control algorithms allowing for real-time response to feedback on droplet formation. This allowed the ML to modify flow rates and other factors in response to environmental changes to achieve uniform microdroplet fabrication at nearly 6 drops per second, a degree of control and throughput that would be unattainable without ML-based automation.

5.2 Polymers for organic photovoltaics

The performance of organic photovoltaic (OPV) devices is determined by properties such as film thickness, blend ratio, solvent environment, and addition of additives or dopants. This creates a large, high-dimensional design space that can be experimentally intractable with conventional Edisonian experimentation. Large design spaces are further confounded by complex, often non-linear interactions that make deriving structure–processing relationships from one-factor-at-a-time experiments challenging or impossible.²²⁰ To overcome this challenge, Harillo-Baños *et al.* came up with a blade-coating HT platform that deposits solution on the substrate while creating a gradient with varying coating speed and solution



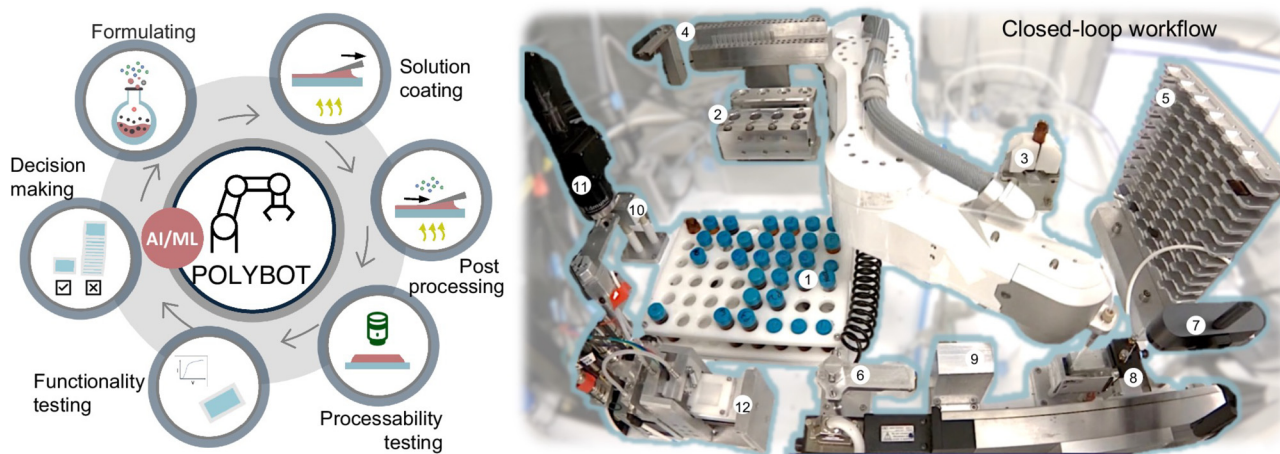


Fig. 8 Demonstration of robotic arm integration in HT synthesis and fabrication of electronic thin films with ML-assisted workflow. Adapted from ref. 24 with permission from Springer Nature, copyright 2025, licensed under CC-BY 4.0.

flow. So, instead of making and testing hundreds of separate OPV devices, a single substrate is coated with thickness gradients. The HT platform also changes ink formulations, solvent mixtures, and polymer : acceptor ratios to allow combinatorial design space exploration. This platform has demonstrated the ability to process over 600 devices across dozens of processing conditions while using less than 50 mg of material.²²¹ Manually fabricating so many devices would be extremely time-consuming and potentially error-prone, but with their HT method they were able to rapidly traverse the design space. Notably, not all HT techniques require ML – more traditional statistical tests can also be effective. In this case, they used one-way ANOVA to determine that solvent choice was the dominant factor affecting device efficiency. They produced blade-coated OPVs with efficiency ranging from 0.08% to 6.43%, with the best candidate demonstrating efficiency up to 14% when optimally fabricated, approaching state-of-the-art at time of publishing.²²² Finally, their large parameter space allowed the creation of performance maps in Hansen solubility space, which they identified as being an effective and computationally inexpensive way of identifying good solvents for similar systems.

In another gradient-based OPV study, roll-to-roll slot-die coating for even more continuous deposition and therefore more HT sampling was used to generate more than 2200 OPVs from *in situ* blend formulations over a large range of film thickness, acceptor : acceptor, and donor : acceptor ratios. The ML models trained on the resulting dataset found high-performance compositions which had OPVs with 10.2% power conversion efficiency.²²³ Moreover, by integrating ML in-line with automation, SDL workflows with BO have demonstrated the ability to sample over 2000 different quaternary active layers in 7 days with significantly reduced material consumption (Fig. 9).²²⁰ These studies highlight the power of HT techniques for the discovery and optimization of polymer electronics.

5.3 Sustainable polymers

Design and synthesis of new polymer materials is slow and resource-intensive due to its reliance on heuristics and chemical intuition, complex iterative synthesis and characterization, and an expansive design space.³² The same is true for polymers designed with sustainability or recyclability as a primary goal. Therefore, Fransen *et al.* developed an HT biodegradation screening platform to synthesize a library of polyesters and polycarbonates with the goal of determining design motifs that impart or improve degradability.⁸³ Using their HT platform they synthesized hundreds of polymers in parallel using a range of HT synthetic schemes including melt polycondensation, melt condensation, and interfacial polymerization. To evaluate biodegradability, they implemented a clear-zone assay to monitor bacterial degradation over time after treating with *Pseudomonas lemoignei*. The researchers used automated optical imaging to measure degradation for each polymer sample over 13 days. Instead of doing hundreds of individual experiments, their use of automated optical imaging helped measure biodegradation of many samples simultaneously. The resulting dataset was used to train ML models that achieved biodegradability prediction accuracy above 82% based on structural descriptors. From their analysis, they found polymers with shorter aliphatic backbones and minimal aromatic content degrade more rapidly while aromatic groups generally suppressed biodegradability. *ortho*- and *para*-substituted benzene rings showed higher degradability compared to *meta*-substitution and backbone ether groups.⁸³

Similarly, rational design of biodegradable PLA derivatives has been accelerated through the use of ML-assisted characterization and design optimization methods,²²⁴ where BO can be used to drastically improve search efficiency. Biodegradation of polymers has also been studied in other contexts, with an ML model from Lin *et al.* reaching predictive R^2 of 0.66 for the degradability of polymers in aqueous environments.⁷¹



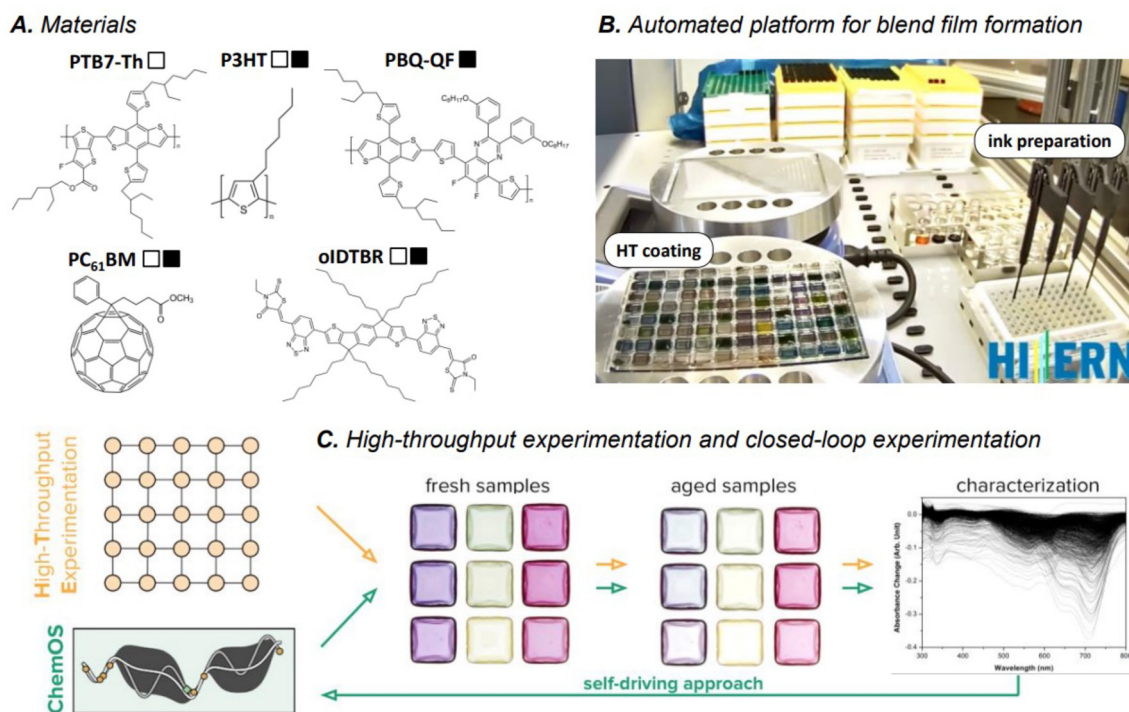


Fig. 9 Self-driving lab for the fabrication of polymer thin films. Reproduced from ref. 220 with permission from Wiley-VCH GmbH, copyright 2020, licensed under CC-BY 4.0.

5.4 Polymers for structural applications

Hu *et al.* employed an ML-assisted approach to design epoxy thermosets that possess high modulus, high tensile strength, and high toughness at the same time.²²⁵ The workflow incorporated transfer learning to leverage information from small molecules and included crosslinking descriptors drawn from gelation theory. The predictive model combined an augmented graph convolutional representation with GPR and was used to virtually screen nearly 250 000 epoxy candidates. Hu and colleagues then synthesized a top-ranked resin that showed superior modulus and elongation compared with existing materials. This integration of modeling and focused validation is important for structural thermosets used in coatings, aerospace composites, and electronics where brittle failure often limits adoption of the material. Their method demonstrated how ML can narrow the search space to a small set of high-value formulations that are likely to meet stringent mechanical specifications.

In another example, Jain *et al.* applied an AL framework to acrylate photopolymers for additive manufacturing.²⁹ The study explored a ternary monomer space that combined rigid monomers, elastomeric monomers, and a crosslinker. The authors used GPR with a two-stage hierarchical model to predict Young's modulus, tensile strength, ultimate strain, and hardness. They then used noisy expected hypervolume improvement, a multi-objective extension of the expected improvement acquisition function commonly used in BO, to recommend the next experimental batch. Each cycle involved automated resin preparation, UV curing, and mechanical testing, forming a closed-

loop. The system rapidly converged on modulus values within 10% of the targets. The selected formulations were validated by fabricating a multimaterial printed structure that combined stiff regions and soft regions, shown in Fig. 10. This work highlights the feasibility of HT/ML for tuning mechanical windows of photocurable systems that are important in aerospace fixtures, automotive parts, and custom medical devices.

5.5 Polymer solubility prediction for synthesis and processing optimization

The accurate prediction of polymer solubility is central to predicting miscibility in polymer blends, membrane science, plastics recycling, solution-based film fabrication, drug delivery, and tuning phase behavior in block copolymer systems.²²⁶ Historically, semi-empirical models based on classical thermodynamic principles, such as the Hildebrand solubility parameter (δ) and Hansen solubility parameters (δ_d , δ_p , δ_h), have been employed to investigate polymer solubility.^{227,228} Solubility parameters can then be determined and tabulated for each component, enabling quick look-ups for solution design. However, their simplicity belies the complexity of solution thermodynamics, so their predictive power is limited. Meanwhile, more detail can be obtained from solution-specific parameters such as the Flory–Huggins interaction parameter χ and, more generally still, activity coefficients (γ) or virial coefficients ($B_{ij}^{(2)}$, $B_{ijk}^{(3)}$, ...), yet the specificity of these quantities makes them of limited use in predicting solubility between generic components.^{229,230} Conventional methods for studying



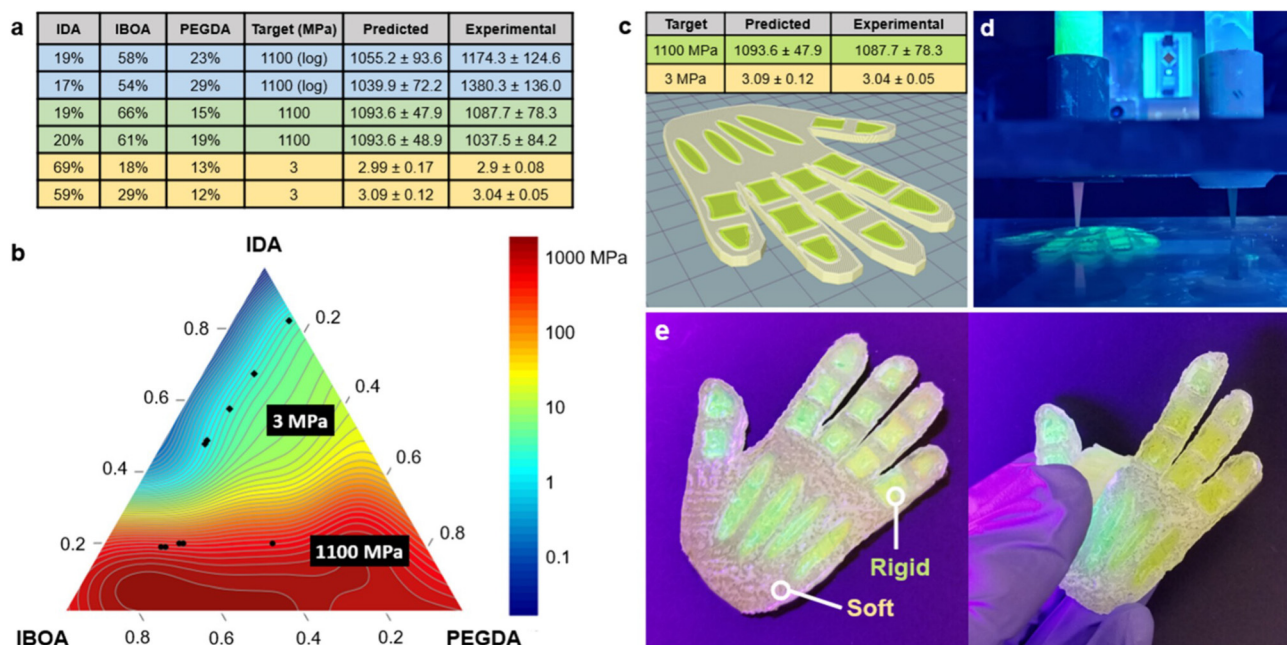


Fig. 10 AL can quickly converge on target properties like Young's modulus values of photopolymers in this multi-material print. Reproduced from ref. 29 with permission from the American Chemical Society, copyright 2024, licensed under CC-BY 4.0.

polymer solubility thus have substantial limitations, including dependence on simplified models and parameters that make assumptions about ideal mixing, negligible entropic effects, and specific intermolecular interactions. In addition, these techniques are associated with significant time and labor investments and are prone to instrument calibrations, experimental limitations, and measurement artifacts.^{231,232} Recently, ML methods have been applied to model polymer-solvent behavior, predict polymer solubility, and discover new polymer-solvent systems *via* inverse design.^{30,203,228,233,234}

In one example, Amrihesari *et al.* used the Crystal16 parallel crystallizer, a “medium-throughput” automated system, to measure polymer solubility.²⁰³ The authors generated a high-quality dataset using turbidity measurements of 30 polymers and 45 solvents across multiple concentrations and temperatures. ML models, including random forest, neural networks, and extreme gradient boosting (XGBoost) regressors, were trained on this dataset to predict transmission-temperature profiles focused on the transmission values during the cooling phase, indicative of solubility behavior. Among the tested approaches, XGBoost showed the best overall performance, achieving an R^2 of 0.98 and RMSE of 6% for predicting percent transmission. Notably, the study extended solubility prediction beyond binary classifications (*i.e.*, soluble/insoluble) by incorporating a “partially soluble” category, thereby improving the practical relevance of predictions for industrial applications such as paint and coating formulation, membrane production, and pharmaceutical development. Although the partial solubility category proved more challenging due to data scarcity, targeted data augmentation significantly improved predictive reliability. This work showed how carefully curated experi-

mental data, when coupled with ML, can accelerate solubility mapping, reduce trial-and-error experimentation, and provide actionable guidance for formulation and process design.

A study by Kern *et al.* addressed a long-standing bottleneck in polymer processing – choosing a solvent that dissolves a given polymer at room temperature by replacing limited Hildebrand/Hansen heuristics with data-driven prediction.²⁰² The authors assembled a curated dataset (3373 polymers, 51 solvents; 11 913 soluble and 8843 insoluble pairs) plus an external set (2909 polymers, 7 solvents) and incorporated structural fingerprints for both polymers and solvents in an attempt to get models to learn physicochemical relationships that might improve generalization beyond the fixed list of solvents used for model training. This research work benchmarked a deep neural network (SolNet2) and a random forest classifier using F1 score under three cross-validation regimes: random, polymer-split (unseen polymers), and solvent-split (unseen solvents). Random forests consistently outperformed the neural net, especially on solvent-split. For their dataset, the incorporation of structural fingerprints did not significantly improve model generalization to unseen solvents. Using Uniform Manifold Approximation and Projection (UMAP) for dimensionality reduction and leave-one-out analysis they identified the modest generalization as an issue of limited solvent diversity. On the hold-out set, class imbalance and novel chemistries further stressed the models, reinforcing the need for broader, more diverse solvent data and for recording experimental factors (molecular weight, concentration, temperature, partial solubility). Overall, the study established a transparent baseline, offered generalizable solvent fingerprints, and demonstrated a robust ML approach that can be progressively enhanced



through the expansion of community-contributed datasets. Further expanding and diversifying polymer–solvent datasets and integrating HT techniques with advanced, interpretable ML models can accelerate polymer solubility prediction and the development of advanced polymeric systems.

5.6 Showcase studies summary

Across the showcase studies highlighted, some trends and observations emerge. While full automation and closed-loop workflows are aspirational, no example herein is truly human-free. At a minimum, every study presented requires the manual preparation of stock solutions, and of course the identification of a problem, identifying and obtaining suitable materials, and setting up a robot with the required experimental equipment. Even within a closed-loop workflow, the selection of preliminary samples upon which the model builds its basis for learning can influence the rate of optimization and final modeling outcomes, so setting an automated system loose with no preliminary instructions is still not likely to produce ideal results.¹⁹⁷

The most common HT equipment from these studies are liquid handlers. Given the complexities of accurately size reducing, measuring, and conveying solid reagents, automated equipment for the preparation of solutions with solid reagents remains out of reach for most laboratories. Therefore, liquid-handling robots, pumps, and automated dispensers including spin coaters and 3D printers remain the most common means of accessing HT data acquisition in a typical laboratory setting.

Further, only a couple of studies employ HT techniques throughout the entirety of the workflow. While we have presented many examples of HT synthesis, purification, and characterization techniques; the relatively nascent stage of this research area means many techniques are created or employed in isolation. For example, all OPV papers referenced can perform HT sample creation, but none demonstrated HT substrate cleaning and preparation. It is not uncommon for only one component of synthesis or characterization to be fully HT at this point in time.

Data is, of course, a motivator for HT methods, and is a driver of improved performance for ML models. With only one exception, the highlighted papers utilize datasets with data points on the order of 10^3 or greater, the largest dataset highlighted in these studies having data points on the order of 10^5 . It is common to supplement experimental data with published datasets when the data is available, a practice demonstrated in about half of these studies. Interestingly, most of these papers do not expressly state throughput, even in papers where HT is the selling point. Quantification of throughput, when available, tends to be vague and abstracted, e.g. “almost 2100 [samples]..within 7 days”.²²⁰ Otherwise, the judgement of “high throughput” tends to rely on implied domain knowledge (a reasonable rate using non-HT methods) or on sheer magnitude – 100 000 data points necessitates some type of accelerated data acquisition in any system. Several additional examples from literature are shown in Table 1 to demonstrate how reporting can vary within literature even for comparable systems. Without a standardized way to report throughput, it is extremely difficult to compare systems and define benchmarks for considering processes as HT. Therefore, we propose that throughput should be reported in samples per hour. By adopting a standard reporting method, researchers can easily compare throughput across systems.

From a modeling standpoint, by far the most common models are the random forest, Gaussian process, and convolutional neural network architectures, with one or more of these or one of their variants appearing in nearly every study highlighted here. Cross-validation and its variants are also common across many of these studies, though error metrics for both training and evaluation remain varied based on the system under exploration.

HT/ML can be daunting to adopt initially, and the ML space is evolving at an extremely rapid pace. However, trends across these studies suggest that incorporation of even some HT methods for repeatable tasks, combined with databases and utilizing even a small number of relatively simple, well-documented models is an extremely viable way of accelerating

Table 1 Reported throughput for various high throughput systems

| No. | Systems | Throughput |
|-----|---|--|
| 1 | PET-RAFT polymerization in multiwell plate | 96 reactions ran in parallel using well plates at a reaction volume of 200 μL . ²³⁵ 384 reactions in parallel for polymerization using 384-wells at a reaction volume of 40 μL . ²³⁶ |
| 2 | Automated continuous flow platform with inline NMR & online SEC/GPC | Traditional offline GPC systems take 20–40 minutes for analysis, reduced to about 12 minutes per sample. ²³⁷ A benchtop NMR spectrometer coupled to the outlet of the flow reactor measures spectra every 15 s. ²³⁸ Automated continuous-flow photopolymerization platform with inline NMR monitoring can collect 120 datapoints in less than 2 hours. ⁹² |
| 3 | Automated polymer synthesis | Systems allow parallelized synthesis of a wide range of polymers with up to 192 reactions. ⁷⁴ 392 polymers synthesized and analyzed in under 40 hours using a 96-well plate workflow. ⁹⁸ |
| 4 | Liquid handling high throughput system | Liquid handling system reduces >80% of the time required for reagent dispensing and allows for a fully automated platform for combinatorial polymer chemistry. ⁸⁹ 672 unique volume transfers usually requires 3 hours of intensive effort. Hamilton MLSTARlet can do that in 30 min. ⁷⁴ |



scientific discovery, and that the barrier to entry may not be as high as it appears at first glance.

Readers interested in practical considerations for setting up and operating SDLs are directed to several recent papers. Day *et al.* discuss high-level practical considerations for HT workflows including the fundamentals of experimental design and design space evaluation for HT experiments, as well as synthetic approaches and characterization and analysis pipelines for HT generation of polymer libraries.⁹⁷ Maroulis *et al.* provide detailed instructions and software for the build and operation of 2 different liquid handlers each costing around \$1000. They also cover some useful fundamentals for active learning approaches, and provide validation of their lab-made automated systems.²³⁹

6 Outlook and perspectives

The synergistic integration of HT/ML has significantly accelerated the pace of polymeric materials design and development in recent years. Recent developments across the HT/ML workflow as highlighted and discussed above offer substantial promise for a future where research and development in applied polymers is systematically supported by automated instrumentation and computational analysis. With labor-intensive synthesis and characterization tasks delegated to automated devices and time-consuming data analysis accomplished by ML algorithms, researchers are freed to invest their time and efforts more effectively in hypothesis generation and

mechanistic investigation. Despite the impressive and encouraging progress, significant challenges and corresponding opportunities for improvement remain.

Data quality and database design are crucial for successful ML model training, since ML models are only as credible as the data they learn from. This is an outstanding issue for polymer informatics, since ML studies based solely on pre-existing datasets suffer from issues of dataset size, quality, and labeling, furthering the need for workflows that can generate high volumes of data with consistent quality, controlled error, and consistent labels for data and metadata.^{17,18,20–22} Further, consideration should be given to the storage, labeling, and sharing of negative data as well – since the lack of publications within “unsuccessful” experimental space can lead to re-treading the same ground or biasing of downstream ML models.²⁴⁰ At minimum, every data point should contain: (1) sample ID with unambiguous polymer representation, (2) complete process record from synthesis to purification, and processing history, and (3) measurements with raw data files, equipment calibration metadata, replicate count, and uncertainty level. Standardization of data recording is necessary for elevating data quality from “best practice” to infrastructure. Well-defined vocabularies and machine-readable schemas should define everything required for data entry. We created a polymer sample record checklist with essential information for data recording critical to ML studies, shown in Fig. 11. We also provide an example of a polymer sample record using our checklist in the SI. However, the investment required for full standardization is unrealistic for individual groups, therefore the field

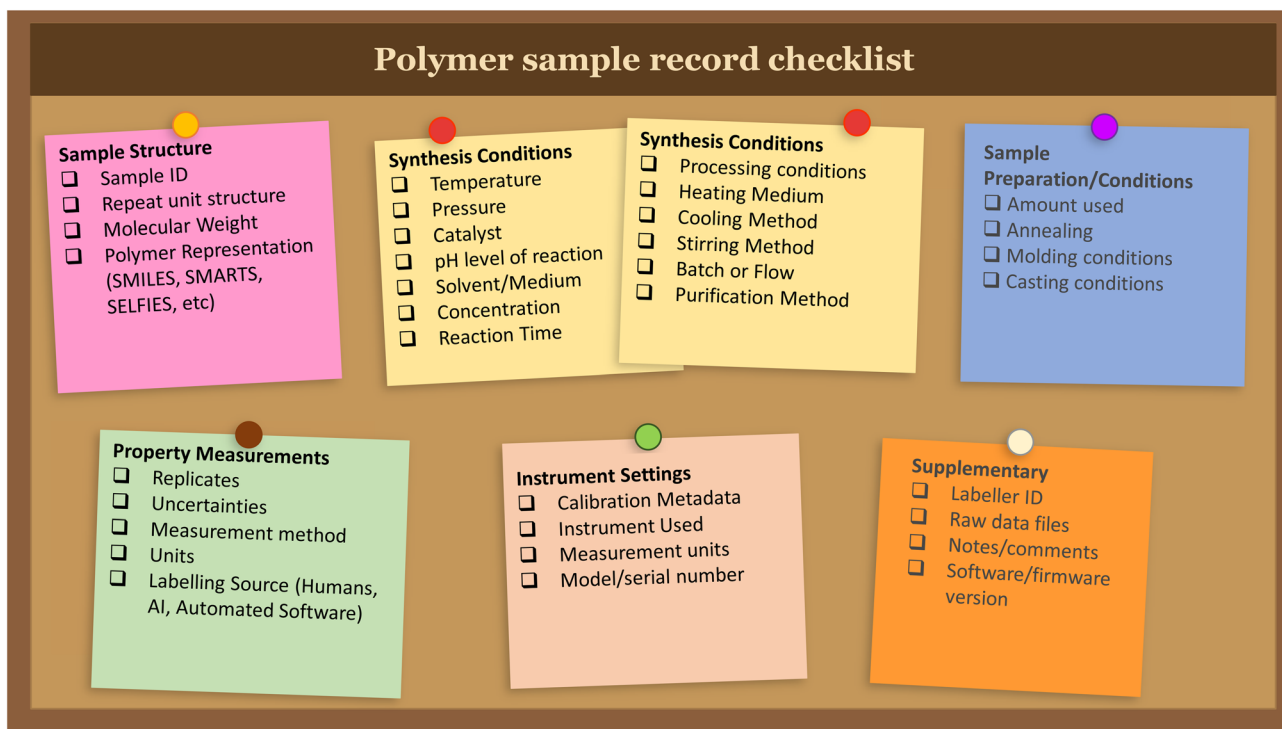


Fig. 11 Abbreviated polymer sample record checklist – some data that should be recorded for every polymer sample (see SI for complete list and example).



needs centralized efforts from organizations such as International Union of Pure and Applied Chemistry (IUPAC) and chemical societies in developing community-maintained ontologies, validators, and ingest pipelines and frameworks. Other disciplines have shown the payoff of such efforts. For example, Crystallographic Information File (CIF),²⁴¹ developed by the International Union of Crystallography, is a machine-readable text format for representing crystallographic information. CIF enables interoperability across software and archives, and tools like checkCIF support automated validation.²⁴²

For HT hardware, the ecosystem spans fully integrated robotic platforms to do-it-yourself (DIY) systems, with corresponding trade-offs in capability, flexibility, and cost. At the high end, installations that combine automated liquid-handling, sample storage, and in-line characterization typically reach the million-dollar scale, offering reliability, vendor support, and safety features but often locking users into proprietary software and constrained workflows. Mid-range, modular systems allow users to customize liquid handlers, imaging stations, and basic analytical tools à la carte, balancing cost against configurability. Entry-level commercial liquid handlers capable of basic dispensing and simple workflows are available on the order of ~\$15 000 at time of publication, lowering the barrier for adoption in individual labs. At the other extreme, open-source designs and 3D-printed parts enable DIY construction of custom liquid-handling and peripheral modules (e.g., camera, sample grabber, etc.) at a fraction of commercial costs, with the added benefit of full control over software and code. These DIY routes, however, transfer the burden from capital expenditure to personnel time and technical expertise in mechanical assembly, electronics, and software integration, and they introduce variability in reliability, safety, and documentation. Across all price-points, lab-scale automation still has major problems with solid-handling and working with heterogeneous mixtures, in addition to lacking dexterity required for complex transfers, sample preparation, and other common experimental routines.²⁴⁰

Given this landscape, field-level coordination led by IUPAC and chemical societies such as the Royal Society of Chemistry, American Chemical Society, and Chinese Chemical Society is essential. Repositories developed and maintained by the research community for hardware designs, materials checklists, calibration procedures, and control software are desired. Analogous to open libraries for 3D-print files, such repositories would accelerate reproducibility and reduce duplicated engineering effort. Community-level demonstration²³⁷ and documentation²³⁹ efforts can further lower the barrier. Because not all individual research groups can afford high-end platforms, shared user facilities, national laboratories, and so-called “cloud labs”^{243,244} are well-positioned to host complex HT experiments instrumentation with trained staff. These facilities have numerous demonstrations of HT/ML workflows,^{245–248} and are actively developing tools and workflows to make HT/ML experimentation even more accessible²⁴⁹ in addition to developing open-source HT tools for use by the community.^{250,251} Improving access to these facilities will

enable researchers to test-run on local, lower-cost systems and execute demanding tasks on centralized infrastructure. This hybrid model, where a local modular system is used for prototyping work, with community standards for interoperability and centralized facilities for complex projects, offers a pragmatic path toward broadening access and reducing the time from concept to results.

Experimental designs should prioritize sampling strategies and allocate for replicates to ensure reproducibility and efficiency within database and hardware constraints. HT experiments can generate data points rapidly, but each data point is associated with a real cost. In our anecdotal experience, a typical polymer sample, from synthesis, processing, to characterization, often costs tens of dollars. Scaling to 1000 samples, the total cost rapidly reaches tens of thousands of dollars. Consideration should therefore be made to cataloging and sharing of both positive and negative data²⁴⁰ to reduce net costs while accelerating output. Additionally, studies should remain cognizant that ML models inherently struggle with out-of-distribution predictions (*i.e.* extrapolation) when selecting experimental designs or ML datasets.^{158,252} In parallel, AL approaches should be leveraged to assist in decision-making and selecting the next experiments. Deliberate planning of experiments, combined with AL, will shorten the polymers design cycle, leading to more cost-efficient information gain and faster convergence to targeted materials properties and performance.

In sum, the promise of HT/ML for applied polymers will be strengthened by a research community that builds together. Shared standards, open-source infrastructure, and collaborative user facilities can turn isolated knowledge gain into an accessible and interoperable knowledge base. As datasets and code become truly reusable, the field can explore mechanisms and principles that unify across chemistries and scales. With an aligned purpose for sharing, HT/ML can mature into an integrative, powerful engine for polymer design and innovation, benefiting society at large, from precision healthcare and environmental remediation to sustainable, circular materials economies.

7 Glossary

This glossary is intended to provide accessible definitions of machine learning and high throughput experimentation terms for readers without formal training in these areas.

Acquisition function: a function used in active learning to determine which data point or experiment should be selected for evaluation next.

Active learning: a machine learning strategy where the model selects the most promising new data points to acquire next. This can be based on learning the experimental space as efficiently as possible, optimizing toward a target, or balancing those objectives simultaneously.

Architecture: the structural design of a model, defining how components are arranged and connected. For example, the architecture of a neural network is described in terms of how many layers and how many nodes per layer it has, while the



architecture of a random forest is described in terms of how many trees with how many branches and of what depth it has.

Automated: part or parts of the process are performed automatically by robotics or computer-driven systems. This can include autosamplers, automated GPC, or other systems that can perform a repetitive, predetermined task.

Autonomous: refers to systems that perform an automated task in response to some input. For example, an instrument with in-line monitoring that can adjust its settings to maintain a target output or property.

Bayesian optimization: an active learning strategy for iterative modeling that uses past results to efficiently select promising new combinations. BO can be used for hyperparameter optimization (as opposed to random or grid search), or for performing active learning over experimental variables.

Classification: a model that assigns inputs to discrete, predefined categories or classes. For example, soluble vs. insoluble (binary classification), block copolymer morphology (multi-class classification).

Closed-loop workflow: a workflow where modeling outputs, especially from an active learning model, are fed back as inputs to guide subsequent actions and experiments.

Clustering: a model that groups data points based on similarity without predefined labels, often used for identifying higher-order relationships between data points, and frequently paired with dimensionality-reduction methods for visualization.

Coefficient of determination (R^2): a regression performance metric that quantifies the variation in the observed data that is explained by the model predictions.

Convolutional neural network: a type of neural network designed to learn spatial or local patterns in structured data, commonly used for images.

Cross-validation: a general framework for estimating model performance by repeatedly training and evaluating a model on different data splits.

Design space: the range of input variables or experimental conditions over which a system is explored.

Data cleaning: the process of identifying and correcting errors, inconsistencies, or missing values in a dataset.

Data scaling: rescaling numerical features to comparable ranges (*e.g.*, normalization or standardization) to improve model performance. Some model architectures like neural networks are sensitive to magnitude and features like molecular weight will dominate model predictions over features like concentration unless scaling is applied.

Data transforming: applying mathematical operations to data (*e.g.*, logarithms, encodings) to improve interpretability or modeling behavior.

Data splitting: dividing data into subsets (*e.g.* training, validation, test) for model training and evaluation.

Descriptor: a numerical feature that encodes some property of a material, molecule, or system for use as an input. ML models do not take inputs like text describing monomer composition, so the numerical representation of that text would be

the monomer's descriptor. Sometimes used synonymously with "feature".

Expected improvement: an acquisition function for active learning that selects new data points for evaluation based on the expected gain over the current best observed result.

Fingerprint: a fixed-length numerical vector that encodes structural or compositional information, commonly used for molecules or polymers.

F1 score: a classification performance metric that is the harmonic mean of precision (true positives/[true positives + false positives]) and recall (true positives/[true positives + false negatives]).

Generalization: a model's ability to make accurate predictions on previously unseen data. How useful a model trained on a specific set of data can also be on related but different data.

Gaussian process regression: a non-parametric regression method that models predictions as probability distributions. Used especially in active learning due to inherently providing both predictions and uncertainty estimates.

Graph neural network: a neural network architecture that learns spatial or local patterns in graph-structured data by propagating information between connected nodes, commonly used for chemical structures.

Grid search: a hyperparameter optimization strategy that exhaustively evaluates combinations of predefined parameter values. More systematic than random search but may miss optimal combinations that were not predefined.

High-dimensional: describes data with a large number of features relative to the number of observations, often complicating analysis and modeling since ML models can easily overfit to noise or redundant signals. Modeling dimensions can be thought of as synonymous with inputs, *e.g.*, a model that predicts elastic modulus based on molecular weight, polymer identity, density, and temperature is 4-dimensional.

Holdout set: a portion of data reserved only for final model evaluation; see Test set.

Hyperparameter: a model setting chosen before training (*e.g.*, learning rate, kernel type, number of nodes and layers) that controls model behavior but is not learned from the data.

k-Fold cross-validation: a cross-validation method where data are split in k subsets ($2 \leq k \leq N$ with N being the total number of data points); each subset is used once as validation while the others are used for training. See also cross-validation.

Leave-one-out analysis: a form of k -fold cross-validation where k is equal to the number of data points.

Loss function: a quantitative measure of model error used for training. Common examples include mean absolute error and root mean squared error (for regression), and log loss (classification).

Machine learning: a class of computational methods that "learn" patterns from data to make predictions or decisions without being explicitly programmed with rules.

Mean absolute error: a regression loss function that is the arithmetic mean of the differences between observed values and predictions. MAE excels at simplicity, interpretability, and explainability.



Model: a mathematical object that maps input data (features/variables) to outputs (predictions) based on learned parameters.

Modular platform: an experimental system composed of interchangeable, automated/automatable components that can be independently modified or reconfigured based on need.

Neural network: a machine learning model composed of interconnected layers of computational nodes that learn relationships between inputs and outputs.

Non-parametric model: models that do not assume a fixed functional form (*e.g.*, linear or exponential) and whose complexity can grow with the amount of available data. These models are often treated as “black boxes” because their internal behavior is not easily interpretable.

One-way ANOVA (analysis of variance): a statistical test used to determine whether the means of three or more groups differ significantly based on a single input variable.

Orthogonality: the degree to which variables vary independently of one another, such that the effect of each variable can be distinguished.

Polymer representation: a numerical or symbolic encoding of polymer chemistry and structure (*e.g.*, SMILES string) used as an input to computational models. Descriptors and fingerprints are derived from polymer representations.

Random forest: a machine learning method that combines predictions from many decision trees trained on different subsets of the data. The ensemble of trees over a single, complex decision tree tends to improve accuracy and robustness.

Random search: a hyperparameter optimization strategy that samples parameter values randomly from defined ranges. Less structured than grid search but stochasticity may allow it to find unintuitive combinations missed by grid search.

Regression: a model that predicts a continuous numerical value. For example, glass transition temperature, modulus, percent conversion.

Roll-to-roll: a continuous manufacturing process in which material is processed as it moves from one roll to another, enabling high throughput and scalable production.

Root mean squared error: a regression loss function that is the quadratic mean of the differences between observed values and predictions. High sensitivity to outliers makes RMSE useful when punishing outliers is a higher priority.

Self-driving: refers to systems that have automated and autonomous elements, but that also engage in decision-making. That is, autonomous and closed-loop systems.

Stratification: a data-splitting strategy that preserves the distribution of key variables (*e.g.*, class labels) across subsets. For example, a solubility classification model trained on 90% insoluble examples is unlikely to perform well on a test set that contains 90% soluble examples.

Test set: a subset of data not used during model training or tuning, used solely to assess final model performance.

Training set: the subset of data used to fit model parameters, what the model “learns” patterns from.

Uniform manifold approximation and projection (UMAP): a dimensionality reduction technique that projects high-dimen-

sional data into a lower-dimensional space while attempting to preserve local structure from the high-dimensional space.

XGBoost (extreme gradient boosting): a machine learning method that uses gradient boosting to build an ensemble of decision trees sequentially, potentially improving prediction performance.

Author contributions

Mashrafee Aryan: conceptualization, writing – original draft, writing – review & editing, visualization. Daniel Struble: conceptualization, writing – original draft, writing – review & editing, visualization. Felix Campbell: writing – original draft, visualization. Saroj Upreti: writing – original draft. S. M. Ashik Abedin: writing – original draft. Aahil Khambawla: writing – original draft. Jeetain Mittal: writing – review & editing, funding acquisition. M. S. Dimitriyev: writing – review & editing, funding acquisition. E. B. Pentzer: writing – review & editing, funding acquisition. S. A. Sukhishvili: writing – review & editing, funding acquisition. Xiaodan Gu: writing – review & editing, funding acquisition. Boran Ma: conceptualization, writing – original draft, writing – review & editing, supervision, funding acquisition.

Conflicts of interest

There are no conflicts to declare.

Data availability

No primary research results, software or code have been included and no new data were generated or analysed as part of this review.

Supplementary information (SI) is available. Polymer sample record checklist and example. See DOI: <https://doi.org/10.1039/d5lp00380f>.

Acknowledgements

The authors thank NSF Division of Materials Research (award numbers 2522693 and 2522694) and DOE Synthesis and Processing Science Program (award number DE-SC0024432) for supporting this work. F. C. thanks NSF REU program (award number 2348780). The authors thank Betty Duggan for her assistance with graphics design. Some elements used in the table of content figure are from flaticon.com.

References

- 1 Y. Zheng, Z. Yu, S. Zhang, X. Kong, W. Michaels, W. Wang, G. Chen, D. Liu, J.-C. Lai, N. Prine, W. Zhang, S. Nikzad, C. B. Cooper, D. Zhong, J. Mun, Z. Zhang,



- J. Kang, J. B.-H. Tok, I. McCulloch, J. Qin, X. Gu and Z. Bao, *Nat. Commun.*, 2021, **12**, 5701.
- 2 L. Ding, Z.-D. Yu, X.-Y. Wang, Z.-F. Yao, Y. Lu, C.-Y. Yang, J.-Y. Wang and J. Pei, *Chem. Rev.*, 2023, **123**, 7421–7497.
- 3 P. E. Jankoski, A.-R. Masoud, J. Dennis, S. Trinh, L. R. DiMartino, J. Shrestha, L. Marrero, J. Hobden, J. Carter, J. Schoen, H. Phelan, A. A. Smith and T. D. Clemons, *Biomacromolecules*, 2025, **26**, 5471–5482.
- 4 S. Chaterji, I. K. Kwon and K. Park, *Prog. Polym. Sci.*, 2007, **32**, 1083–1122.
- 5 Y. Guan, K. P. Meyers, S. K. Mendon, G. Hao, J. R. Douglas, S. Trigwell, S. I. Nazarenko, D. L. Patton and J. W. Rawlins, *ACS Appl. Mater. Interfaces*, 2016, **8**, 33210–33220.
- 6 A. Kafi, H. Wu, J. Langston, O. Atak, H. Kim, S. Kim, W. P. Fahy, R. Reber, J. Misasi, S. Bateman and J. H. Koo, *J. Appl. Polym. Sci.*, 2020, **137**, 49117.
- 7 M. Meirbekov, A. Kuandyk, M. Sadykov, M. Nurzhanov, N. Yesbolov, B. Baiserikov, I. Ablakatov, L. Mustafa, B. Medyanova, A. Kulbekov, S. Orazbek and A. Yermekov, *Polymers*, 2025, **17**, 1419.
- 8 J. Shen, J. Liang, X. Lin, H. Lin, J. Yu and Z. Yang, *Int. J. Polym. Sci.*, 2020, **2020**, 8838160.
- 9 A. M. Virshup, J. Contreras-García, P. Wipf, W. Yang and D. N. Beratan, *J. Am. Chem. Soc.*, 2013, **135**, 7296–7303.
- 10 D. C. Struble, B. G. Lamb and B. Ma, *MRS Commun.*, 2024, **14**, 752–770.
- 11 T.-S. Lin, C. W. Coley, H. Mochigase, H. K. Beech, W. Wang, Z. Wang, E. Woods, S. L. Craig, J. A. Johnson, J. A. Kalow, K. F. Jensen and B. D. Olsen, *ACS Cent. Sci.*, 2019, **5**, 1523–1531.
- 12 L. Schneider, D. Walsh, B. Olsen and J. d. Pablo, *Digital Discovery*, 2024, **3**, 51–61.
- 13 J. Paturej, S. S. Sheiko, S. Panyukov and M. Rubinstein, *Sci. Adv.*, 2016, **2**, e1601478.
- 14 N. Hadjichristidis, H. Iatrou, M. Pitsikalis and J. Mays, *Prog. Polym. Sci.*, 2006, **31**, 1068–1132.
- 15 D. Y. Lee, J. T. Pham, J. Lawrence, C. H. Lee, C. Parkos, T. Emrick and A. J. Crosby, *Adv. Mater.*, 2013, **25**, 1248–1253.
- 16 N. D. Ogbonna, M. Dearman, C.-T. Cho, B. Bharti, A. J. Peters and J. Lawrence, *JACS Au*, 2022, **2**, 898–905.
- 17 J.-P. Correa-Baena, K. Hippalgaonkar, J. van Duren, S. Jaffer, V. R. Chandrasekhar, V. Stevanovic, C. Wadia, S. Guha and T. Buonassisi, *Joule*, 2018, **2**, 1410–1420.
- 18 E. O. Pyzer-Knapp, J. W. Pitera, P. W. J. Staar, S. Takeda, T. Laino, D. P. Sanders, J. Sexton, J. R. Smith and A. Curioni, *NPJ Comput. Mater.*, 2022, **8**, 84.
- 19 I. H. Sarker, *SN Comput. Sci.*, 2021, **2**, 160.
- 20 L. Himanen, A. Geurts, A. S. Foster and P. Rinke, *Adv. Sci.*, 2019, **6**, 1900808.
- 21 A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. A. Persson, *APL Mater.*, 2013, **1**, 011002.
- 22 B. Ma, N. J. Finan, D. Jany, M. E. Deagen, L. S. Schadler and L. C. Brinson, *Macromolecules*, 2023, **56**, 3945–3953.
- 23 S. Oliver, L. Zhao, A. J. Gormley, R. Chapman and C. Boyer, *Macromolecules*, 2019, **52**, 3–23.
- 24 C. Wang, Y.-J. Kim, A. Vriza, R. Batra, A. Baskaran, N. Shan, N. Li, P. Darancet, L. Ward, Y. Liu, M. K. Y. Chan, S. K. R. S. Sankaranarayanan, H. C. Fry, C. S. Miller, H. Chan and J. Xu, *Nat. Commun.*, 2025, **16**, 1498.
- 25 B. Dadfar, B. Alemdag and G. Kabay, *Macromol. Rapid Commun.*, 2025, e00380.
- 26 M. Abolhasani and E. Kumacheva, *Nat. Synth.*, 2023, **2**, 483–492.
- 27 P. A. Beaucage, D. R. Sutherland and T. B. Martin, *Macromolecules*, 2024, **57**, 8661–8670.
- 28 A. Vriza, H. Chan and J. Xu, *Chem. Mater.*, 2023, **35**, 3046–3056.
- 29 A. Jain, C. D. Armstrong, V. R. Joseph, R. Ramprasad and H. J. Qi, *ACS Appl. Mater. Interfaces*, 2024, **16**, 17992–18000.
- 30 C. D. Stubbs, Y. Kim, E. C. Quinn, R. Pérez-Soto, E. Y.-X. Chen and S. Kim, *Digital Discovery*, 2025, **4**, 424–437.
- 31 J.-M. Lu, J.-Z. Pan, Y.-M. Mo and Q. Fang, *Artif. Intell. Chem.*, 2024, **2**, 100057.
- 32 S. Baudis and M. Behl, *Macromol. Rapid Commun.*, 2022, **43**, 2100400.
- 33 R. Nsouli, G. Galiyan and L. K. G. Ackerman-Biegasiewicz, *Angew. Chem., Int. Ed.*, 2025, **64**, e202506588.
- 34 L. Yu, B. Chen, Z. Li, Y. Su, X. Jiang, Z. Han, Y. Zhou, D. Yan, X. Zhu and R. Dong, *Giant*, 2024, **18**, 100252.
- 35 L. Ma, W. Li, J. Yuan, J. Zhu, Y. Wu, H. He and X. Pan, *Macromol. Rapid Commun.*, 2025, e00361.
- 36 W. Ge, R. De Silva, Y. Fan, S. A. Sisson and M. H. Stenzel, *Adv. Mater.*, 2025, **37**, 2413695.
- 37 R. Upadhyaya, A. Punia, M. J. Kanagala, L. Liu, M. Lamm, T. A. Rhodes and A. J. Gormley, *ACS Appl. Polym. Mater.*, 2021, **3**, 1525–1536.
- 38 R. A. Patel and M. A. Webb, *ACS Appl. Bio Mater.*, 2024, **7**, 510–527.
- 39 X. Rodríguez-Martínez, E. Pascual-San-José and M. Campoy-Quiles, *Energy Environ. Sci.*, 2021, **14**, 3301–3322.
- 40 Y. Dai, H. Chan, A. Vriza, F. Kim, Y. Wang, W. Liu, N. Shan, J. Xu, M. Weires, Y. Wu, Z. Cao, C. S. Miller, R. Divan, X. Gu, C. Zhu, S. Wang and J. Xu, Adaptive AI decision interface for autonomous electronic material discovery, *arXiv*, 2025, preprint, arXiv:2504.13344 [cond-mat], <https://arxiv.org/abs/2504.13344>.
- 41 Y. Wang, R. Sriramoju, S. Upreti, K. H. Chan, M. Noack, E. Schaible, D. English, A. Hexemer, W. Koepp, C. Zhu and X. Gu, *ChemRxiv*, 2026, preprint, <https://chemrxiv.org/doi/10.26434/chemrxiv-2026-z7g5t>.
- 42 M. Avital-Shmilovici, X. Liu, T. Shaler, A. Lowenthal, P. Bourbon, J. Snider, A. Tambo-Ong, C. Repellin, K. Yniguez, L. Sambucetti, P. B. Madrid and N. Collins, *ACS Cent. Sci.*, 2022, **8**, 86–101.
- 43 A. Round, F. Felisaz, L. Fodinger, A. Gobbo, J. Huet, C. Villard, C. E. Blanchet, P. Pernot, S. McSweeney, M. Roessle, D. I. Svergun and F. Cipriani, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2015, **71**, 67–75.



- 44 T. Schuett, P. Endres, T. Standau, S. Zechel, R. Q. Albuquerque, C. Brütting, H. Ruckdäschel and U. S. Schubert, *Adv. Funct. Mater.*, 2024, **34**, 2309844.
- 45 D. T. McQuade and P. H. Seeberger, *J. Org. Chem.*, 2013, **78**, 6384–6389.
- 46 M. B. Plutschack, B. Pieber, K. Gilmore and P. H. Seeberger, *Chem. Rev.*, 2017, **117**, 11796–11893.
- 47 R. C. K. Montalbo, M.-J. Wu and H.-L. Tu, *RSC Adv.*, 2024, **14**, 11258–11265.
- 48 S. G. Newman and K. F. Jensen, *Green Chem.*, 2013, **15**, 1456–1472.
- 49 F. M. Akwi and P. Watts, *Chem. Commun.*, 2018, **54**, 13894–13928.
- 50 R. A. Meyers, *Handbook of Petrochemicals Production Processes*, McGraw-Hill Education, 2nd edn, 2019.
- 51 T. E. Glier, M. Vakili and M. Trebbin, *J. Polym. Res.*, 2020, **27**, 333.
- 52 A. Sivokhin, D. Orekhov, O. Kazantsev, K. Otopkova, O. Sivokhina, I. Chuzhaykin, A. Ovchinnikov, O. Zamyshlyayeva, I. Pavlova, O. Ozhogina and M. Chubenko, *Polymers*, 2024, **16**, 134.
- 53 A. K. Padmakumar, N. K. Singha, M. Ashokkumar, F. A. Leibfarth and G. G. Qiao, *Macromolecules*, 2023, **56**, 6920–6927.
- 54 A. Nagaki and Y. Ashikari, *Polym. J.*, 2025, **57**, 143–148.
- 55 Z. Jin, H. Wang, X. Hu, Y. Liu, Y. Hu, S. Zhao, N. Zhu, Z. Fang and K. Guo, *React. Chem. Eng.*, 2022, **7**, 1026–1036.
- 56 Y. Takahashi and A. Nagaki, *Molecules*, 2019, **24**, 1532.
- 57 C. L. G. I. Davidson, M. E. Lott, L. Trachsel, A. J. Wong, R. A. Olson, D. I. Pedro, W. G. Sawyer and B. S. Sumerlin, *ACS Macro Lett.*, 2023, **12**, 1224–1230.
- 58 S. M. Giannitelli, E. Limiti, P. Mozetic, F. Pinelli, X. Han, F. Abbruzzese, F. Basoli, D. D. Rio, S. Scialla, F. Rossi, M. Trombetta, L. Rosanò, G. Gigli, Z. J. Zhang, E. Mauri and A. Rainer, *Nanoscale*, 2022, **14**, 11415–11428.
- 59 M. Reis, F. Gusev, N. G. Taylor, S. H. Chung, M. D. Verber, Y. Z. Lee, O. Isayev and F. A. Leibfarth, *J. Am. Chem. Soc.*, 2021, **143**, 17677–17689.
- 60 Y. Wang, D. Struble, S. Upreti, Z. Xie, K. H. Chan, Y. Liu, C. Zhu, P. Ashby, W. Xia, D. Patton, B. Ma and X. Gu, Automated and High-Throughput Phase Separation Control for Supramolecular Polymer Blends Enabled by Machine Learning, *ChemRxiv*, 2025preprint, <https://chemrxiv.org/engage/chemrxiv/article-details/690ae086ef936fb4a2676662>.
- 61 T. Ishiyama, Y. Kobayashi, H. Nakamura, M. Aizawa, K. Hisano, S. Kubo and A. Shishido, *Macromolecules*, 2024, **57**, 7430–7438.
- 62 S. Camarri, A. Mariotti, C. Galletti, E. Brunazzi, R. Mauri and M. V. Salvetti, *Ind. Eng. Chem. Res.*, 2020, **59**, 3669–3686.
- 63 K. Huanbutta, P. Sriamornsak, K. Suwanpitak, N. Klinchuen, T. Deebgkum, V. Teppitak and T. Sangnim, *Int. J. Nanomed.*, 2023, **18**, 7889–7900.
- 64 Z. Whiteley, H. M. K. Ho, Y. X. Gan, L. Panariello, G. Gkogkos, A. Gavriilidis and D. Q. M. Craig, *Nanoscale Adv.*, 2021, **3**, 2039–2055.
- 65 T. A. Bauer, J. Schramm, F. Fenaroli, S. Siemer, C. I. Seidl, C. Rosenauer, R. Bleul, R. H. Stauber, K. Koynov, M. Maskos and M. Barz, *Adv. Mater.*, 2023, **35**, 2210704.
- 66 M. W. Losey, M. A. Schmidt and K. F. Jensen, *Ind. Eng. Chem. Res.*, 2001, **40**, 2555–2562.
- 67 N. Chanchaona, L. Ding, S. Lin, S. Sarwar, S. Dimartino, A. J. Fletcher, D. M. Dawson, K. Konstas, M. R. Hill and C. H. Lau, *J. Mater. Chem. A*, 2023, **11**, 9859–9867.
- 68 Z. Dong, Z. Wen, F. Zhao, S. Kuhn and T. Noël, *Chem. Eng. Sci.: X*, 2021, **10**, 100097.
- 69 S. Basak and A. Bandyopadhyay, *ACS Appl. Eng. Mater.*, 2024, **2**, 1190–1208.
- 70 Y. Yu, S.-H. Shen, J.-B. Chen, Q.-R. Ding, H.-X. Zhang and J. Zhang, *J. Am. Chem. Soc.*, 2025, **147**, 20770–20777.
- 71 C. Lin and H. Zhang, *Environ. Sci. Technol.*, 2025, **59**, 1253–1263.
- 72 G. Carretero, H. K. Samarasekara, A. Battigelli and B. Mojsoska, *Small*, 2025, **21**, 2406128.
- 73 A. Kayser, K. Klein, D. Babushkina, A. Sakse, G. M. Kenne, U. I. Gerling-Driessen, M. Tabatabai, M. Epple and L. Hartmann, *Chem. – Eur. J.*, 2025, **31**, e202500497.
- 74 M. Tamasi, S. Kosuri, J. DiStefano, R. Chapman and A. J. Gormley, *Adv. Intell. Syst.*, 2020, **2**, 1900126.
- 75 Y. Asano, K. Okada, S. Nakagawa, N. Yoshie and J. Shiomi, *Rob. Auton. Syst.*, 2025, **185**, 104868.
- 76 Y. Wu, A. Vriza, D. Ozgulbas, R. Vescovi, J. Zhou, Z. Wang, S. Hu, Y. Zhang, Q. Yang, A. Österholm, J. Reynolds, S. Sankaranarayanan, M. Chan, I. Foster, H. Chan, J. Mei and J. Xu, Autonomous Synthesis and Inverse Design of Electronic Polymers with High Efficiency and Accuracy *ChemRxiv*, 2025preprint, <https://chemrxiv.org/engage/chemrxiv/article-details/6806c318927d1c2e66%207ec5>.
- 77 T. Schuett, J. Kimmig, S. Zechel and U. S. Schubert, *Polymers*, 2022, **14**, 292.
- 78 F. Kong, L. Yuan, Y. F. Zheng and W. Chen, *SLAS Technol.*, 2012, **17**, 169–185.
- 79 V. F. Jafari, Z. Mossayebi, S. Allison-Logan, S. Shabani and G. G. Qiao, *Chem. – Eur. J.*, 2023, **29**, e202301767.
- 80 J. R. Shuluk, C. D. Wight, J. R. Howard, M. E. King, S. R. Moor, R. J. DeHoog, S. D. Dahlhauser, L. S. Eberlin and E. V. Anslyn, *JACS Au*, 2025, **5**, 1232–1242.
- 81 J. J. Green, R. Langer and D. G. Anderson, *Acc. Chem. Res.*, 2008, **41**, 749–759.
- 82 L. K. Petersen, A. V. Chavez-Santoscoy and B. Narasimhan, *J. Visualized Exp.*, 2012, 3882.
- 83 K. A. Fransen, S. H. M. Av-Ron, T. R. Buchanan, D. J. Walsh, D. T. Rota, L. Van Note and B. D. Olsen, *Proc. Natl. Acad. Sci. U. S. A.*, 2023, **120**, e2220021120.
- 84 C. Avila, C. Cassani, T. Kogej, J. Mazuela, S. Sarda, A. D. Clayton, M. Kossenjans, C. P. Green and R. A. Bourne, *Chem. Sci.*, 2022, **13**, 12087–12099.
- 85 J. Wu, X. Yang, Y. Pan, T. Zuo, Z. Ning, C. Li and Z. Zhang, *J. Flow Chem.*, 2023, **13**, 385–404.
- 86 I. Terzioglu, C. Ventura-Hunter, J. Ulbrich, E. Saldivar-Guerra, U. S. Schubert and C. Guerrero-Sanchez, *Polymers*, 2022, **14**, 4835.



- 87 T. Schuett, J. Kimmig, S. Zechel and U. S. Schubert, *Polymers*, 2020, **12**, 2095.
- 88 B. Lin, J. L. Hedrick, N. H. Park and R. M. Waymouth, *J. Am. Chem. Soc.*, 2019, **141**, 8921–8927.
- 89 J. Lee, P. Mulay, M. J. Tamasi, J. Yeow, M. M. Stevens and A. J. Gormley, *Digital Discovery*, 2023, **2**, 219–233.
- 90 P. Q. Velasco, K. Y. A. Low, C. J. Leong, W. T. Ng, S. Qiu, S. Jhunjhunwala, B. Li, A. Qian, K. Hippalgaonkar and J. J. W. Cheng, *Digital Discovery*, 2024, **3**, 1011–1020.
- 91 N. Yoshikawa, Y. Asano, D. N. Futaba, K. Harada, T. Hitosugi, G. N. Kanda, S. Matsuda, Y. Nagata, K. Nagato, M. Naito, T. Natsume, K. Nishio, K. Ono, H. Ozaki, W. Shin, J. Shiomi, K. Shizume, K. Takahashi, S. Takeda, I. Takeuchi, R. Tamura, K. Tsuda and Y. Ushiku, *Digital Discovery*, 2025, **4**, 1384–1403.
- 92 G. D. Ammini, L. J. Weerathna and T. Junkers, *Chem.: Methods*, 2025, **5**, e202500025.
- 93 K. Verstraete, A.-L. Buckinx, N. Zaquen and T. Junkers, *Macromolecules*, 2021, **54**, 3865–3872.
- 94 J. P. Bhimani, R. Ouseph and R. A. Ward, *Nephrol., Dial., Transplant.*, 2010, **25**, 3990–3995.
- 95 L. Brocken, P. D. Price, J. Whittaker and I. R. Baxendale, *React. Chem. Eng.*, 2017, **2**, 656–661.
- 96 R. Upadhya, M. J. Kanagala and A. J. Gormley, *Macromol. Rapid Commun.*, 2019, **40**, 1900528.
- 97 E. C. Day, S. S. Chittari, M. P. Bogen and A. S. Knight, *ACS Polym. Au*, 2023, **3**, 406–427.
- 98 C. Stubbs, T. Congdon, J. Davis, D. Lester, S.-J. Richards and M. I. Gibson, *Macromolecules*, 2019, **52**, 7603–7612.
- 99 G. Wu, H. Zhou, J. Zhang, Z.-Y. Tian, X. Liu, S. Wang, C. W. Coley and H. Lu, *Nat. Synth.*, 2023, **2**, 515–526.
- 100 J. Chen, V. Bhat and C. J. Hawker, *J. Am. Chem. Soc.*, 2024, **146**, 8650–8658.
- 101 J. M. Lee, J. Kwon, S. J. Lee, H. Jang, D. Kim, J. Song and K. T. Kim, *Sci. Adv.*, 2022, **8**, eabl8614.
- 102 M. Christensen, I. Chiciudean, P. Jablonski, A.-M. Tanase, V. Shapaval and H. Hansen, *PLoS One*, 2023, **18**, e0282623.
- 103 A. Paul, L. Wander, R. Becker, C. Goedecke and U. Braun, *Environ. Sci. Pollut. Res.*, 2019, **26**, 7364–7374.
- 104 S. Shiwani, I. Latka, J. Popp, C. Krafft and I. W. Schie, *ACS Omega*, 2025, **10**, 33675–33688.
- 105 O. Nassar, M. Jouda, M. Rapp, D. Mager, J. G. Korvink and N. MacKinnon, *Microsyst. Nanoeng.*, 2021, **7**, 30.
- 106 W. Pointer, R. Radmall, O. Tooley, J. Town, D. C. J. Haggart, Z. Zhai, X. Yang, D. W. Lester, P. Wilson and D. M. Haddleton, *Polym. Chem.*, 2025, **16**, 3329–3343.
- 107 E. A. Murphy, Y.-Q. Chen, K. Albanese, J. R. Blankenship, A. Abdilla, M. W. Bates, C. Zhang, C. M. Bates and C. J. Hawker, *Macromolecules*, 2022, **55**, 8875–8882.
- 108 E. A. Murphy, C. Zhang, C. M. Bates and C. J. Hawker, *Acc. Chem. Res.*, 2024, **57**, 1202–1213.
- 109 E. A. Murphy, K. G. Roth, M. W. Bates, M. C. Murphy, J. Edmund, C. M. Bates and C. J. Hawker, *Macromolecules*, 2025, **58**, 8369–8376.
- 110 B. Lamb, S. Upreti, Y. Wang, D. Struble, C. Zhu, G. Freychet, X. Gu and B. Ma, Machine Learning Framework for Characterizing Processing-Structure Relationship in Block Copolymer Thin Films, *arXiv*, 2025, preprint, arXiv:2505.23064 [cond-mat], <https://arxiv.org/abs/2505.23064>.
- 111 M. J. Werny, K. B. Siebers, N. H. Friederichs, C. Hendriksen, F. Meirer and B. M. Weckhuysen, *J. Am. Chem. Soc.*, 2022, **144**, 21287–21294.
- 112 Y. Luo, M. Gu, C. E. R. Edwards, M. T. Valentine and M. E. Helgeson, *Soft Matter*, 2022, **18**, 3063–3075.
- 113 J. C. Meredith, A. Karim and E. J. Amis, *Macromolecules*, 2000, **33**, 5760–5762.
- 114 I. Biran, L. Houben, A. Kossoy and B. Rybitchinski, *J. Phys. Chem. C*, 2024, **128**, 5988–5995.
- 115 S. Benaglia, C. A. Amo and R. Garcia, *Nanoscale*, 2019, **11**, 15289–15297.
- 116 J. Ren and Q. Zou, *Beilstein J. Nanotechnol.*, 2017, **8**, 1563–1570.
- 117 S. Benaglia, V. G. Gisbert, A. P. Perrino, C. A. Amo and R. Garcia, *Nat. Protoc.*, 2018, **13**, 2890–2907.
- 118 B. Alldritt, P. Hapala, N. Oinonen, F. Urtev, O. Krejci, F. F. Canova, J. Kannala, F. Schulz, P. Liljeroth and A. S. Foster, *Sci. Adv.*, 2020, **6**, eaay6913.
- 119 A. Chandrashekar, P. Belardinelli, M. A. Bessa, U. Staufer and F. Alijani, *Nanoscale Adv.*, 2022, **4**, 2134–2143.
- 120 M. Neuenschwander, S. H. Andany, M. Kangül, N. Hosseini and G. E. Fantner, 2021 21st International Conference on Solid-State Sensors, Actuators and Microsystems (Transducers), 2021, pp. 22–25.
- 121 A. Glia, M. Deliorman and M. A. Qasaimeh, *Adv. Sci.*, 2022, **9**, 2201489.
- 122 N. P. Cowieson, C. J. C. Edwards-Gayle, K. Inoue, N. S. Khunti, J. Douth, E. Williams, S. Daniels, G. Preece, N. A. Krumpa, J. P. Sutter, M. D. Tully, N. J. Terrill and R. P. Rambo, *J. Synchrotron Radiat.*, 2020, **27**, 1438–1446.
- 123 A. Yaghmur and I. Hamad, *Molecules*, 2022, **27**, 4602.
- 124 K. Tang, A. Shaw, S. Upreti, H. Zhao, Y. Wang, G. T. Mason, J. Aguinaga, K. Guo, D. Patton, D. Baran, S. Rondeau-Gagné and X. Gu, *Chem. Mater.*, 2025, **37**, 756–765.
- 125 P. S. Rahimabadi, M. Khodaei and K. R. Koswattage, *X-Ray Spectrom.*, 2020, **49**, 348–373.
- 126 E. Rossi, J. M. Wheeler and M. Sebastiani, *Curr. Opin. Solid State Mater. Sci.*, 2023, **27**, 101107.
- 127 A. B. Irez, J. Hay, I. Miskioglu and E. Bayraktar, *Mechanics of Composite and Multi-functional Materials*, Cham, 2018, vol. 6, pp. 1–9.
- 128 T. Oellers, V. G. Arigela, C. Kirchlechner, G. Dehm and A. Ludwig, *ACS Comb. Sci.*, 2020, **22**, 142–149.
- 129 M. Fischer, J.-J. Wiegand and I. Kuehnert, *AIP Conf. Proc.*, 2024, **3158**, 130005.
- 130 A. R. Piacenti, C. Adam, N. Hawkins, R. Wagner, J. Seifert, Y. Taniguchi, R. Proksch and S. Contera, *Macromolecules*, 2024, **57**, 1118–1127.
- 131 J. Zhang, Y. Liu, D. Chandra, P. Sekhar, M. Singh, Y. Tong, E. Kucukdeger, H. Y. Yoon, A. P. Haring, M. Roman, Z. J. Kong and B. N. Johnson, *Appl. Mater. Today*, 2023, **30**, 101720.



- 132 J. E. Griffith, Y. Chen, Q. Liu, Q. Wang, J. J. Richards, D. Tullman-Ercek, K. R. Shull and M. Wang, *Mater. Horiz.*, 2023, **10**, 97–106.
- 133 P. Salas-Ambrosio, C. I. Gupit, J. M. Urueña, Y. Luo, J. M. Hankett, R. Gupta, M. T. Valentine, H. D. Maynard and M. E. Helgeson, *Polym. Chem.*, 2024, **15**, 1758–1766.
- 134 D. Mangal, A. Jha, D. Dabiri and S. Jamali, *Curr. Opin. Colloid Interface Sci.*, 2025, **75**, 101873.
- 135 T. G. Fox Jr. and P. J. Flory, *J. Appl. Phys.*, 1950, **21**, 581–591.
- 136 J. Peng, E. Jury, P. Dönnies and C. Ciurtin, *Front. Pharmacol.*, 2021, **12**, 720694.
- 137 M. Ceriotti, C. Clementi and O. A. von Lilienfeld, *Chem. Rev.*, 2021, **121**, 9719–9721.
- 138 A. M. Schweidtmann, E. Esche, A. Fischer, M. Kloft, J.-U. Repke, S. Sager and A. Mitsos, *Chem. Ing. Tech.*, 2021, **93**, 2029–2039.
- 139 D. Lafuente, B. Cohen, G. Fiorini, A. A. García, M. Bringas, E. Morzan and D. Onna, *J. Chem. Educ.*, 2021, **98**, 2892–2898.
- 140 A. Jayaraman and B. Olsen, *Macromolecules*, 2024, **57**, 7685–7688.
- 141 E. Bhardwaj, H. Gujral, S. Wu, C. Zogheib, T. Maharaj and C. Becker, The 2024 ACM Conference on Fairness, Accountability, and Transparency, 2024, pp. 1055–1067.
- 142 J. Sun, W. Zhang, Y. Chen, B. B. Hoar, H. Sheng, J. Y. Yang, Q. Gu and C. Liu, *J. Phys. Chem. C*, 2025, **129**, 1044–1051.
- 143 *Data Preprocessing in Python*, <https://www.geeksforgeeks.org/machine-learning/data-preprocessing-machine-learning-python/>, Section: Machine Learning.
- 144 *Introduction to RDKit—Python for Data Science in Chemistry*, https://education.molssi.org/python-data-science-chemistry/rdkit_descriptors/rdkit.html.
- 145 D. Rogers and M. Hahn, *J. Chem. Inf. Model.*, 2010, **50**, 742–754.
- 146 *Splitting Data for Machine Learning Models*, <https://www.geeksforgeeks.org/machine-learning/splitting-data-for-machine-learning-models/>, Section: Machine Learning.
- 147 E. Jain, J. Neeraja, B. Banerjee and P. Ghosh, A Diagnostic Approach to Assess the Quality of Data Splitting in Machine Learning, *arXiv*, 2022, preprint, arXiv:2206.11721 [stat], <https://arxiv.org/abs/2206.11721>.
- 148 J. Brownlee, *A Gentle Introduction to k-fold Cross-Validation*, 2018, <https://www.machinelearningmastery.com/k-fold-cross-validation/>.
- 149 L. Franceschi, M. Donini, V. Perrone, A. Klein, C. Archambeau, M. Seeger, M. Pontil and P. Frasconi, *Hyperparameter Optimization in Machine Learning*, *arXiv*, 2025, preprint, arXiv:2410.22854 [stat], <https://arxiv.org/abs/2410.22854>.
- 150 O. Rainio, J. Teuho and R. Klén, *Sci. Rep.*, 2024, **14**, 6086.
- 151 A. Bajaj, *Performance Metrics in Machine Learning [Complete Guide]*, 2022, <https://neptune.ai/blog/performance-metrics-in-machine-learning-complete-guide>.
- 152 T. A. Meyer, C. Ramirez, M. J. Tamasi and A. J. Gormley, *ACS Polym. Au*, 2023, **3**, 141–157.
- 153 L. Tao, V. Varshney and Y. Li, *J. Chem. Inf. Model.*, 2021, **61**, 5395–5413.
- 154 Y. Zhao, N. Schiffmann, A. Koeppel, N. Brandt, E. C. Bucharsky, K. G. Schell, M. Selzer and B. Nestler, *Front. Mater.*, 2022, **9**, 1–12.
- 155 R. A. Patel, C. H. Borca and M. A. Webb, *Mol. Syst. Des. Eng.*, 2022, **7**, 661–676.
- 156 P. Karande, B. Gallagher and T. Y.-J. Han, *Chem. Mater.*, 2022, **34**, 7650–7665.
- 157 A. Zekić, *Computation*, 2025, **13**, 169.
- 158 E. R. Antoniuk, S. Zaman, T. Ben-Nun, P. Li, J. Diffenderfer, B. Sahin, O. Smolenski, T. Hsu, A. M. Hiszpanski, K. Chiu, B. Kaillkhura and B. V. Essen, BOOM: Benchmarking Out-Of-distribution Molecular Property Predictions of Machine Learning Models, *arXiv*, 2025, preprint, arXiv:2505.01912 [cs], <https://arxiv.org/abs/2505.01912>.
- 159 X. Zhong, B. Gallagher, S. Liu, B. Kaillkhura, A. Hiszpanski and T. Y.-J. Han, *NPJ Comput. Mater.*, 2022, **8**, 204.
- 160 V. Belle and I. Papantonis, *Front. Big Data*, 2021, **4**, 688969.
- 161 S. Lundberg and S.-I. Lee, A Unified Approach to Interpreting Model Predictions, *arXiv*, 2017, preprint, arXiv:1705.07874 [cs], <https://arxiv.org/abs/1705.07874>.
- 162 W. Sha, Y. Li, S. Tang, J. Tian, Y. Zhao, Y. Guo, W. Zhang, X. Zhang, S. Lu, Y.-C. Cao and S. Cheng, *InfoMat*, 2021, **3**, 353–361.
- 163 T. B. Martin and D. J. Audus, *ACS Polym. Au*, 2023, **3**, 239–258.
- 164 C. Yan and G. Li, *Adv. Intell. Syst.*, 2023, **5**, 2200243.
- 165 M. Fang, S. Tang, Z. Fan, Y. Shi, N. Xu and Y. He, *J. Phys. Chem. A*, 2024, **128**, 2286–2294.
- 166 C. Chen, R. Liang, S. Xia, D. Hou, B. Abdoulaye, J. Tao, B. Yan, Z. Cheng and G. Chen, *Fuel*, 2023, **332**, 126177.
- 167 M. Kedzierski, M. Falcou-Préfol, M. E. Kerros, M. Henry, M. L. Pedrotti and S. Bruzaud, *Chemosphere*, 2019, **234**, 242–251.
- 168 X. Yan, Z. Cao, A. Murphy and Y. Qiao, *J. Environ. Chem. Eng.*, 2022, **10**, 108130.
- 169 P. Polyak, P. Chaber, M. Musioł, G. Adamus, M. Kowalczyk, J. E. Puskas and M. El Fray, *Anal. Sci.*, 2025, **41**, 1015–1027.
- 170 S. P. Chelvam, A. J. Y. Ng, J. Huang, E. Lee, M. Baranski, D. Yong, R. B. H. Williams, S. L. Springs and R. J. Ram, *Sci. Rep.*, 2025, **15**, 7631.
- 171 R. Mamede, F. Pereira and J. Aires-de Sousa, *Sci. Rep.*, 2021, **11**, 23720.
- 172 A. D. McNaughton, R. P. Joshi, C. R. Knutson, A. Fnu, K. J. Luebke, J. P. Malerich, P. B. Madrid and N. Kumar, *J. Chem. Inf. Model.*, 2023, **63**, 1462–1471.
- 173 A. U. Hassan and M. J. Aljaafreh, *Coatings*, 2025, **15**, 558.
- 174 PrathamModi, *Convolutional Neural Networks for Dummies*, 2023, <https://medium.com/@prathammodi01/convolutional-neural-networks-for-dummies-a-step-by-step-cnn-tutorial-e68f464d608f>.
- 175 Z. Liang, Z. Tan, R. Hong, W. Ouyang, J. Yuan and C. Zhang, *J. Chem. Inf. Model.*, 2023, **63**, 5971–5980.



- 176 K. Ferji, *Nanoscale*, 2025, **17**, 18777–18786.
- 177 S. V. Kalinin, M. Ziatdinov, J. Hinkle, S. Jesse, A. Ghosh, K. P. Kelley, A. R. Lupini, B. G. Sumpter and R. K. Vasudevan, *ACS Nano*, 2021, **15**, 12604–12627.
- 178 A. Ghosh, B. G. Sumpter, O. Dyck, S. V. Kalinin and M. Ziatdinov, *NPJ Comput. Mater.*, 2021, **7**, 100.
- 179 M. G. Wessels and A. Jayaraman, *ACS Polym. Au*, 2021, **1**, 153–164.
- 180 X. Fang, E. A. Murphy, P. A. Kohl, Y. Li, C. J. Hawker, C. M. Bates and M. Gu, *J. Polym. Sci.*, 2025, **63**, 1433–1440.
- 181 L.-B. Li, *Chin. J. Polym. Sci.*, 2018, **36**, 1093–1102.
- 182 Y. Lu, E. Yang, J. Zhu, S. Liu, K. Cui, H. Guo and L. Li, *Rev. Sci. Instrum.*, 2024, **95**, 093907.
- 183 W. Chen, D. Liu and L. Li, *Polym. Cryst.*, 2019, **2**, 10043.
- 184 Y. Lin, W. Chen, L. Meng, D. Wang and L. Li, *Soft Matter*, 2020, **16**, 3599–3612.
- 185 A. Arbe, F. Alvarez and J. Colmenero, *Polymers*, 2020, **12**, 3067.
- 186 L. V. Tiihonen, M. P. Weir, A. J. Parnell, S. C. Boothroyd, D. W. Johnson, R. M. Dalgliesh, M. Bleuel, C. P. Duif, W. G. Bouwman, R. L. Thompson, K. S. Coleman, N. Clarke, W. A. Hamilton, A. L. Washington and S. R. Parnell, *Soft Matter*, 2024, **20**, 8663–8674.
- 187 D. J. Beltran-Villegas, M. G. Wessels, J. Y. Lee, Y. Song, K. L. Wooley, D. J. Pochan and A. Jayaraman, *J. Am. Chem. Soc.*, 2019, **141**, 14916–14930.
- 188 S. Yu, J. Chen, G. Gomard, H. Hölscher and U. Lemmer, *Adv. Opt. Mater.*, 2023, **11**, 2203134.
- 189 C. M. Wolf, L. Guio, S. Scheiwiller, V. Pakhnyuk, C. Luscombe and L. D. Pozzo, *ACS Polym. Au*, 2021, **1**, 134–152.
- 190 P. C. St John, C. Phillips, T. W. Kemper, A. N. Wilson, Y. Guan, M. F. Crowley, M. R. Nimlos and R. E. Larsen, *J. Chem. Phys.*, 2019, **150**, 234111.
- 191 C. M. Heil, Y. Ma, B. Bharti and A. Jayaraman, *JACS Au*, 2023, **3**, 889–904.
- 192 C. M. Heil, A. Patil, A. Dhinojwala and A. Jayaraman, *ACS Cent. Sci.*, 2022, **8**, 996–1007.
- 193 Z. Ye, Z. Wu and A. Jayaraman, *JACS Au*, 2021, **1**, 1925–1936.
- 194 S. V. R. Akepati, N. Gupta, J. Shah, S. Kronenberger, V. Venkat, R. A. Sridhar, S. Bianco, D. J. Adams and A. Jayaraman, *ACS Meas. Sci. Au*, 2026, **6**, 1–20.
- 195 R. Arboretti, R. Ceccato, L. Pegoraro and L. Salmaso, *Qual. Reliab. Eng. Int.*, 2022, **38**, 3357–3378.
- 196 R. Arboretti, R. Ceccato, L. Pegoraro and L. Salmaso, *Qual. Reliab. Eng. Int.*, 2022, **38**, 1131–1156.
- 197 Q. M. Gallagher and M. A. Webb, *Digital Discovery*, 2025, **4**, 135–148.
- 198 E. Bagci and B. Işik, *Int. J. Adv. Manuf. Tech.*, 2006, **31**, 10–17.
- 199 T. Erzurumlu and H. Oktem, *Mater. Des.*, 2007, **28**, 459–465.
- 200 S. M. Karazi, A. Issa and D. Brabazon, *Opt. Lasers Eng.*, 2009, **47**, 956–964.
- 201 J. E. S. Urrutia, M. A. Villalobos, J. E. Silva Urrutia and M. A. Villalobos, *Poblac. Salud Mesoam.*, 2022, **20**, 1–27.
- 202 J. Kern, S. Venkatram, M. Banerjee, B. Brettmann and R. Ramprasad, *Phys. Chem. Chem. Phys.*, 2022, **24**, 26547–26555.
- 203 M. Amrihesari, J. Kern, H. Present, S. M. Briceno, R. Ramprasad and B. Brettmann, *J. Phys. Chem. B*, 2024, **128**, 12786–12797.
- 204 H. Jung, C. D. Stubbs, S. Kumar, R. Pérez-Soto, S.-m. Song, Y. Kim and S. Kim, *Digital Discovery*, 2025, **4**, 1492–1504.
- 205 J. Pei, C. Cai, X. Zhu, G. Wang and B. Yan, *Adv. Mater. Res.*, 2012, **455–456**, 436–442.
- 206 G. M. Casanola-Martin, A. Karuth, H. Pham-The, H. González-Díaz, D. C. Webster and B. Rasulev, *Commun. Chem.*, 2024, **7**, 226.
- 207 M.-X. Zhu, T. Deng, L. Dong, J.-M. Chen and Z.-M. Dang, *IET Nanodielectr.*, 2022, **5**, 24–38.
- 208 R. Gurnani, D. Kamal, H. Tran, H. Sahu, K. Scharm, U. Ashraf and R. Ramprasad, *Chem. Mater.*, 2021, **33**, 7008–7016.
- 209 L. Chen, C. Kim, R. Batra, J. P. Lightstone, C. Wu, Z. Li, A. A. Deshmukh, Y. Wang, H. D. Tran, P. Vashishta, G. A. Sotzing, Y. Cao and R. Ramprasad, *NPJ Comput. Mater.*, 2020, **6**, 61.
- 210 S. Brierley-Croft, P. D. Olmsted, P. J. Hine, R. J. Mandle, A. Chaplin, J. Grasmeyer and J. Mattsson, *Macromolecules*, 2025, **58**, 6407–6417.
- 211 A. Agnihotri and N. Batra, *Distill*, 2020, **5**, e26.
- 212 J. Wang, *Comput. Sci. Eng.*, 2023, **25**, 4–11.
- 213 A. V. Tobias and A. Wahab, *R. Soc. Open Sci.*, 2025, **12**, 250646.
- 214 S. T. Knox, K. E. Wu, N. Islam, R. O'Connell, P. M. Pittaway, K. E. Chingono, J. Oyekan, G. Panoutsos, T. W. Chamberlain, R. A. Bourne and N. J. Warren, *Polym. Chem.*, 2025, **16**, 1355–1364.
- 215 M. J. Tamasi, R. A. Patel, C. H. Borca, S. Kosuri, H. Mugnier, R. Upadhyaya, N. S. Murthy, M. A. Webb and A. J. Gormley, *Adv. Mater.*, 2022, **34**, 2201809.
- 216 M. M. Noack, P. H. Zwart, D. M. Ushizima, M. Fukuto, K. G. Yager, K. C. Elbert, C. B. Murray, A. Stein, G. S. Doerk, E. H. Tsai, R. Li, G. Freychet, M. Zhernenkov, H.-Y. N. Holman, S. Lee, L. Chen, E. Rotenberg, T. Weber, Y. L. Goc, M. Boehm, P. Steffens, P. Mutti and J. A. Sethian, *Nat. Rev. Phys.*, 2021, **3**, 685–697.
- 217 G. Xu, R. Zhang and T. Luo, Self-Driving Laboratory Optimizes the Lower Critical Solution Temperature of Thermoresponsive Polymers, *arXiv*, 2025, preprint, arXiv:2509.05351 [cond-mat], <https://arxiv.org/abs/2509.05351>.
- 218 J. A. O'Callaghan, N. P. Kamat, K. B. Vargo, R. Chattaraj, D. Lee and D. A. Hammer, *Eur. Phys. J. E*, 2024, **47**, 37.
- 219 S. Shin, O. D. Land, W. D. Seider, J. Lee and D. Lee, *Small*, 2025, **21**, 2412099.
- 220 S. Langner, F. Häse, J. D. Perea, T. Stubhan, J. Hauch, L. M. Roch, T. Heumueller, A. Aspuru-Guzik and C. J. Brabec, Beyond Ternary OPV: High-Throughput Experimentation and Self-Driving Laboratories Optimize



- Multi-Component Systems, *arXiv*, 2019, preprint, arXiv:1909.03511 [physics], <https://arxiv.org/abs/1909.03511>.
- 221 A. Harillo-Baños, Q. Fan, S. Riera-Galindo, E. Wang, O. Inganäs and M. Campoy-Quiles, *ChemSusChem*, 2022, **15**, e202101888.
- 222 Q. Fan, Q. An, Y. Lin, Y. Xia, Q. Li, M. Zhang, W. Su, W. Peng, C. Zhang, F. Liu, L. Hou, W. Zhu, D. Yu, M. Xiao, E. Moons, F. Zhang, T. D. Anthopoulos, O. Inganäs and E. Wang, *Energy Environ. Sci.*, 2020, **13**, 5017–5027.
- 223 N. G. An, J. Y. Kim and D. Vak, *Energy Environ. Sci.*, 2021, **14**, 3438–3446.
- 224 R. Fujita, Y. Amamoto and J. Kikuchi, *Npj Mater. Degrad.*, 2025, **9**, 72.
- 225 Y. Hu, W. Zhao, L. Wang, J. Lin and L. Du, *ACS Appl. Mater. Interfaces*, 2022, **14**, 55004–55016.
- 226 B. A. Miller-Chou and J. L. Koenig, *Prog. Polym. Sci.*, 2003, **28**, 1223–1270.
- 227 M. Rubinstein and R. H. Colby, *Polymer physics*, Oxford Univ. Press, Oxford, repr edn, 2014.
- 228 J. Ethier, E. R. Antoniuk and B. Brettmann, *Soft Matter*, 2024, **20**, 5652–5669.
- 229 G. Zante, *Artif. Intell. Chem.*, 2024, **2**, 100069.
- 230 H.-C. Liao, Y.-H. Lin, C.-H. Peng and Y.-P. Li, *ACS Eng. Au*, 2025, **5**, 530–539.
- 231 D. Medarević, J. Djuriš, P. Barmplexis, K. Kachrimanis and S. Ibrić, *Pharmaceutics*, 2019, **11**, 372.
- 232 A. Zhenova, *Polym. Int.*, 2020, **69**, 895–901.
- 233 R. Hassan and M. R. Kazemi, *Sci. Rep.*, 2025, **15**, 31157.
- 234 M. Amrihesari, M. Banerjee, R. Olmedo and B. Brettmann, *Macromol. Rapid Commun.*, 2025, e00454.
- 235 G. Ng, J. Yeow, R. Chapman, N. Isahak, E. Wolvetang, J. J. Cooper-White and C. Boyer, *Macromolecules*, 2018, **51**, 7600–7607.
- 236 A. J. Gormley, J. Yeow, G. Ng, O. Conway, C. Boyer and R. Chapman, *Angew. Chem., Int. Ed.*, 2018, **57**, 1557–1562.
- 237 J. V. Herck, I. Abeysekera, A.-L. Buckinx, K. Cai, J. Hooker, K. Thakur, E. V. d. Reydt, P.-J. Voortter, D. Wyers and T. Junkers, *Digital Discovery*, 2022, **1**, 519–526.
- 238 M. Rubens, J. Van Herck and T. Junkers, *ACS Macro Lett.*, 2019, **8**, 1437–1441.
- 239 A. Maroulis, D. Waynor, Q. Gallagher, R. Patel, M. Tamasi, D. C. Radford, M. Webb and A. Gormley, A User's Guide to Your First Self-Driving Liquid Handling Lab, *ChemRxiv*, 2025preprint, <https://chemrxiv.org/engage/chemrxiv/article-details/68f6948baecc32c656822ef03>.
- 240 M. Seifrid, R. Pollice, A. Aguilar-Granda, Z. M. Chan, K. Hotta, C. T. Ser, J. Vestfrid, T. C. Wu and A. Aspuru-Guzik, *Acc. Chem. Res.*, 2022, **55**, 2454–2466.
- 241 S. R. Hall, F. H. Allen and I. D. Brown, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 1991, **47**, 655–685.
- 242 A. L. Spek, *Acta Crystallogr., Sect. E: Crystallogr. Commun.*, 2020, **76**, 1–11.
- 243 D. S. Arias and R. E. Taylor, *Adv. Mater. Technol.*, 2024, **9**, 2400084.
- 244 A. M. Bran, S. Cox, O. Schilter, C. Baldassari, A. D. White and P. Schwaller, *Nat. Mach. Intell.*, 2024, **6**, 525–535.
- 245 G. S. Doerk, A. Stein, S. Bae, M. M. Noack, M. Fukuto and K. G. Yager, *Sci. Adv.*, 2023, **9**, eadd3687.
- 246 M. M. Noack, G. S. Doerk, R. Li, J. K. Streit, R. A. Vaia, K. G. Yager and M. Fukuto, *Sci. Rep.*, 2020, **10**, 17663.
- 247 N. J. Szymanski, B. Rendy, Y. Fei, R. E. Kumar, T. He, D. Milsted, M. J. McDermott, M. Gallant, E. D. Cubuk, A. Merchant, H. Kim, A. Jain, C. J. Bartel, K. Persson, Y. Zeng and G. Ceder, *Nature*, 2023, **624**, 86–91.
- 248 P. A. Beaucage and T. B. Martin, *Chem. Mater.*, 2023, **35**, 846–852.
- 249 S. Mathur, N. v. der Vleuten, K. G. Yager and E. H. R. Tsai, *Mach. Learn. Sci. Technol.*, 2025, **6**, 025051.
- 250 G. Hao, E. J. Roberts, T. Chavez, Z. Zhao, E. A. Holman, H. Yanxon, A. Green, H. Krishnan, D. Ushizima, D. McReynolds, N. Schwarz, P. H. Zwart, A. Hexemer and D. Y. Parkinson, IS&T International Symposium on Electronic Imaging, 2023, **35**, IPAS-290.
- 251 L. R. DaCosta, K. Sytwu, C. K. Groschner and M. C. Scott, *NPJ Comput. Mater.*, 2024, **10**, 165.
- 252 J. Hu, D. Liu, N. Fu and R. Dong, *Digital Discovery*, 2024, **3**, 300–312.

