12 June 2025

Autonomous Synthesis and Inverse Design of Electronic Polymers with High Efficiency and Accuracy

Yukun Wu[1,2], Aikaterini Vriza[1], Doga Ozgulbas[1], Rafael Vescovi[1], Jianing Zhou[1,2], Zhiyang Wang[2], Shiyu Hu[1], Yuepeng Zhang[1], Qiaomu Yang[1], Anna Österholm[4], John Reynolds[4], Subramanian Sankaranarayanan[1], Maria Chan[1], Ian Foster[1], Henry Chan[1], Jianguo Mei[2], Jie Xu[1,3]

1. Argonne National Laboratory
2. Purdue University West Lafayette
3. University of Chicago
4. Georgia Institute of Technology

Abstract

The design and synthesis of functional polymers, aimed at targeted properties through specific structures, has long been challenged by their complex and often nonlinear structure-property relationships. Key processes, including knowledge accumulation for predictive design and experimental refinement and validation, are traditionally labor-insensitive and time-consuming, making it difficult to balance accuracy and efficiency. Here, we introduce an accelerated, autonomous system for the on-demand synthesis of electronic polymers that achieves desired electrochromic functionality with high accuracy and efficiency. Our approach leverages large language model-assisted data mining, a physics-informed copolymer machine learning model, and an AI-driven autonomous robotic workflow in the Polybot lab. Within 72 hours, Polybot autonomously synthesized electrochromic polymers with targeted, unreported color values, including green polymers with specific absorption profiles, precisely fine-tuning copolymer structures with a 5% step size in comonomer composition within a three-monomer system. A publicly accessible electrochromic polymer informatics database has also been created to foster knowledge exchange.

Keywords

Electrochromic Polymers, Autononous Chemistry

Autonomous Synthesis and Inverse Design of Electronic Polymers with High Efficiency and Accuracy

Authors: Yukun Wu[1,2]†, Aikaterini Vriza[1]†, Doga Ozgulbas[3], Rafael Vescovi[3], Jianing Zhou[1,2], Zhiyang Wang[2], Shiyu Hu[4], Yuepeng Zhang[4], Qiaomu Yang[1], Anna M. Österholm[5], John R. Reynolds[5], Subramanian K.R.S. Sankaranarayanan[1], Maria K.Y. Chan[1], Ian T. Foster[3], Jianguo Mei[2]*, Henry Chan[1]*, Jie Xu[1,6]*

Affiliations:

[1]Nanoscience and Technology Division, Argonne National Laboratory, Lemont, IL 60439, USA.

[2]Department of Chemistry, Purdue University, West Lafayette, IN 47907, USA.

[3]Data Science and Learning Division, Argonne National Laboratory, Lemont, IL 60439, USA

[4]Applied Materials Division, Argonne National Laboratory, Lemont, IL 60439, USA.

[5]School of Chemistry & Biochemistry, School of Materials Science & Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA

[6]Pritzker School of Molecular Engineering, The University of Chicago, Chicago, Illinois 60637, USA

†These authors contributed equally to this work.

*Corresponding author. Email: xuj@anl.gov; jgmei@purdue.edu; hchan@anl.gov

Abstract:

The design and synthesis of functional polymers, aimed at targeted properties through specific structures, has long been challenged by their complex and often nonlinear structure-property relationships. Key processes, including knowledge accumulation for predictive design and experimental refinement and validation, are traditionally labor-insensitive and time-consuming, making it difficult to balance accuracy and efficiency. Here, we introduce an accelerated, autonomous system for the on-demand synthesis of electronic polymers that achieves desired electrochromic functionality with high accuracy and efficiency. Our approach leverages large language model-assisted data mining, a physics-informed copolymer machine learning model, and an AI-driven autonomous robotic workflow in the Polybot lab. Within 72 hours, Polybot autonomously synthesized electrochromic polymers with targeted, unreported color values, including green polymers with specific absorption profiles, precisely fine-tuning copolymer structures with a 5% step size in comonomer composition within a three-monomer system. A publicly accessible electrochromic polymer informatics database has also been created to foster knowledge exchange.

Main Text:

The synthesis of functional polymers aimed at precise and tailored properties has traditionally been a time-consuming, labor-intensive process, often relying on iterative trial-and-error methods to achieve the desired polymer characteristics. This challenge is particularly pronounced in the development of electronic polymers, where subtle variations in polymer structure can significantly impact their optoelectronic performance[1-3]. Rigorous design and validation procedures are often required to explore complex polymer structure-property relationships and ensure precise functionality. Despite notable advancements in computational tools and modeling techniques,

accurately predicting the intricate relationships between polymer structures, properties, and performance has remained elusive.

The inverse design and synthesis of functional polymers hinges on two critical and time-consuming tasks: the accumulation of knowledge for predictive design and the collection of experimental data for the refinement and validation of predictions. Both tasks must be accomplished while managing the inherent trade-off between accuracy and efficiency. However, in general, knowledge accumulation is a long process that requires an in-depth review of existing literature in the research field and continuous learning through the scientific method, not to mention the importance of essential experimental refinement and validation. Recent progress in integrating artificial intelligence (AI) with automated experimentation has notably accelerated materials research, facilitating the development of functional small organic molecules and inorganic materials with specific properties[4-12]. The incorporation of data mining tools has streamlined the distillation of information from the scientific literature and experimental datasets [13]. Nevertheless, establishing AI-powered automated laboratories for polymer scientists remains challenging due to the complex structure-property relationships and diverse, heterogeneous nature of polymers, which make both AI and automated experimentation design very different from those for inorganic or small organic molecules[14-19.] To effectively use AI-guided automated labs for functional polymers discovery, several major challenges need be addressed: 1) creating comprehensive databases for functional polymers[20]; 2) developing accurate analytical and data-driven methods that can predict the complexity of structure-property relationships, especially in statistical copolymer systems[21,22]; and 3) establishing AI-integrated advanced lab automation technologies for handling complex polymer studies, from synthesis to purification and characterization feedback[23,24]. While a few pioneering initiatives have introduced autonomous

platforms for polymer research, such as formulation or condition optimization, the impact of monomer structural diversity on polymer functionalities remains underexplored, and the complexity of autonomous lab workflow design needs substantial enhancement for a wide range of functional polymer discovery[25-30].

In this context, we present an accelerated, autonomous, and on-demand discovery approach for electronic polymers. This approach achieves high-accuracy electrochromic functionality through large language model (LLM)-assisted data mining, a Set-Transformer-based copolymer machine learning (ML) predictive model, and an AI-driven robotic workflow in the Polybot lab for self-guided, iterative polymer synthesis, purification, processing, and characterization. Electrochromic polymers (ECPs) are known for their ability to switch between colored and bleached states in response to an electrical stimulus, making them highly valuable in energy-efficient technologies including smart windows, displays, and sensors[31,32]. Despite decades of empirical studies providing invaluable insights, achieving precise color accuracy remains a challenging task.[33] It is both labor-intensive and time-consuming, requiring meticulous adjustments and repeated trials to fine-tune copolymer structures and account for the complex factors that influence absorbance and color values. Here, within 72 hours, Polybot autonomously produced ECPs with targeted, previously unreported orange and green color values, which includes the particularly challenging green ECPs that require specific absorption profiles and finely tuned copolymer structures with a 5% step size in comonomer composition within a three-monomer system. The data extracted and produced in this work allows us to establish a publicly accessible interactive ECP informatics and an open database[34], which we envision will foster knowledge exchange and accelerate further developments in the electronic polymer field.

Double-loop inverse design strategy in Polybot

The properties of functional polymers are intricately tied to their structural compositions and conformations. For instance, the color nuances of ECPs depend strongly on the monomer energy levels, dihedral angles between neighboring monomers, monomer ratios in final structures, and polymer chain lengths[35,36]. To synthesize polymers accurately and efficiently with on-demand properties, Polybot adopts a nested "double-loop" inverse design strategy (Fig. 1a). The left-side loop focuses on synthesizing polymers using AI/ML-suggested monomer building blocks, discovering appropriate synthetic conditions to achieve targeted polymer chemical structures, and ensuring the necessary reaction yield for property evaluation in the right-side loop. Following this, the right-side loop explores targeted structures with precise and desirable functionalities, characterizes the functionalities of the synthesized polymer, and develops surrogate ML models to predict the outcomes of unperformed experiments. Ultimately, the targeted properties of the final polymer materials are realized through this comprehensive double-loop inverse design.

This double-loop inverse design strategy is realized in Polybot via a fusion of advanced laboratory automation (Fig. 1b), data analytics infrastructure, and AI/ML design intelligence[19]. The Polybot central computing hub orchestrates the automated execution of polymer synthesis, purification, sample processing, and property evaluation. These complex operations are seamlessly coordinated through a network of cooperative multi-robotic systems, operating under a unified control network (Fig. 1c). A workflow execution interface handles robot execution commands and feedback, as well as data transfer across instruments, ECP database, and AI/ML model[37]. With this integrated modular platform, Polybot synthesized over 90 ECPs with new comonomer compositions and distributions in a week while expanding the color spectrum and demonstrating electrochromic functionality in various devices and aerosol-jet printed flexible electrochromic

displays (Fig. 1d, Supplementary Section 1 and Movie 1). In the following sections, we will describe the innovations in Polybot that enabled this inverse design study.

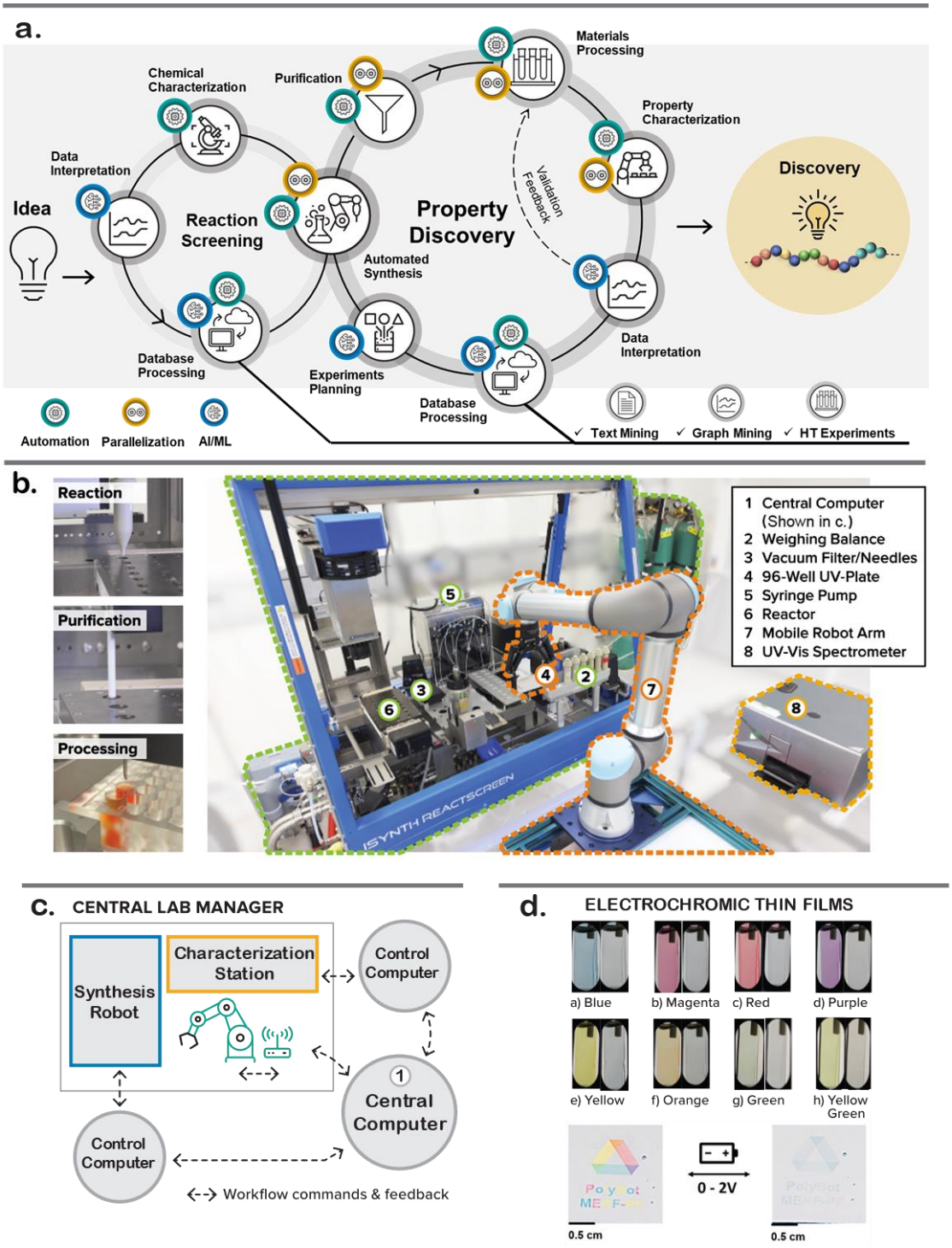

Fig. 1 Overview of the inverse design of ECPs in Polybot. (a) Illustrations of the double-loop inverse design strategy and how each component fits into the double-loop design; Left-side circle: reaction condition screening workflow; Right-side circle: polymer property optimization cycle. **(b)** Left: key operations performed by the synthesis robot. Right: experimental modules for the autonomous discovery of ECPs. **(c)** Lab layout and workflow management: the central computer sends commands to control computers associated with the different automated systems; the automated systems then perform the task and provide feedback via the control computers to the central computer. **(d)**, Top (a-h): randomly selected ECP thin films from different color themes in the colored state (left) and transmissive state (right); Bottom: multi-colored electrochromic display printed through an aerosol-gel printer in the colored (left) and transmissive (right) state.

Automated data mining for predictive design and synthesis conditions optimization

The journey towards inverse design starts with the crucial step of accumulating structure-property knowledge for predictive design, which can be a laborious and time-consuming process. Information about material structures and properties is often scattered throughout scientific literature, appearing in both text and images[33,36]. To mitigate these challenges, we developed a versatile large language model (LLM)-assisted data extraction and organization workflow within our LLM-guided materials data mining pipeline (Fig. 2a) [39]. Unlike experimental screening by a single group with limited resources, this approach expands the design space by leveraging prior literature. While the initial development of this LLM-assisted data-extraction workflow required notable effort, it resulted in a generally applicable, automated framework for extracting knowledge from both text and images, making it broadly useful for various materials systems (Fig. S1), with its first demonstration in this electrochromic polymer system. Specifically, we applied this workflow to extract data in the ECP related literature, focusing on thiophene-based ECPs known for their electrochromic performance (Table S1). These copolymers exhibit characteristics such as a wide color scope, sub-second switching speed, and high contrast over 80%[35,40]. This selection was not intended to limit scope but rather to balance feasibility and effort, resulting in a practical

and well-rounded initial dataset. Our LLM-assisted workflow streamlines the extraction of SMILES strings of repeating units in ECPs from chemical structure images and digitizes ultraviolet-visible (UV-Vis) spectra. To elucidate structure-property relationships, the workflow categorizes ECP structures based on their conjugated backbones and appended sidechains. Furthermore, to minimize inconsistencies in color descriptions, the workflow directly calculates color properties in the CIELab color space from the UV-Vis spectra (Supplementary Section 2.2, Fig. S2) [33,38,41]. A statistical analysis of our ECP database reveals a diverse distribution of monomers, covering a broad spectrum of colors (Fig. 2b, c, Fig. S3). Approximately 70% of these copolymers consist of two types of comonomer units, while the remaining 30% are composed of three types of comonomer units (Fig. S4).

In designing generalizable synthesis conditions for copolymerizing various comonomers into ECPs, the Pd-catalyzed C-C cross-coupling step based on direct C-H arylation stands out due to its atomic economy and broad applicability across diverse monomer types[42]. However, the complexity of reaction parameters and degrees of freedom associated with direct C-H arylation conditions necessitates optimization for efficiency. To systematically refine reaction conditions, we conducted reaction optimization experiments guided by statical insights provided from ML-assisted data mining while incorporating human considerations. We developed a multi-agent LLM assistant system that leverages prompt engineering techniques to employ three sequential LLM assistants, for quickly identifying commonly used catalysts, ligands, bases, and solvents for direct C-H arylation. The first assistant identifies paragraphs in text specifically related to the synthesis method. The second assistant performs named entity recognition to categorize the four components: catalysts, ligands, bases, and solvents used in polymerization. Finally, the third assistant refines the data by eliminating semantic duplicates (e.g., ‘dry THF’ and ‘THF’) and reports occurrences

in all four categories (Fig. 2b, Supplementary Section 2.4). The extracted synthetic knowledge serves as a blueprint, we further refine the optimization space by ensuring experimental feasibility while preventing unnecessary expansion of the search space (refinement rational is provided in Supplementary Section 3.1). The extracted synthetic knowledge serves as a blueprint for formulating synthesis conditions during reaction optimization. By conducting high-throughput reaction screening on a representative model reaction system and analyzing proton nuclear magnetic resonance (^{1}H-NMR) data, we identified optimal conditions that yield high reaction efficiency (Fig. 2d, Supplementary Section 3, Scheme S1, Table S2, and Fig. S5). The highest yielding condition was then employed for copolymerization in subsequent reactions.

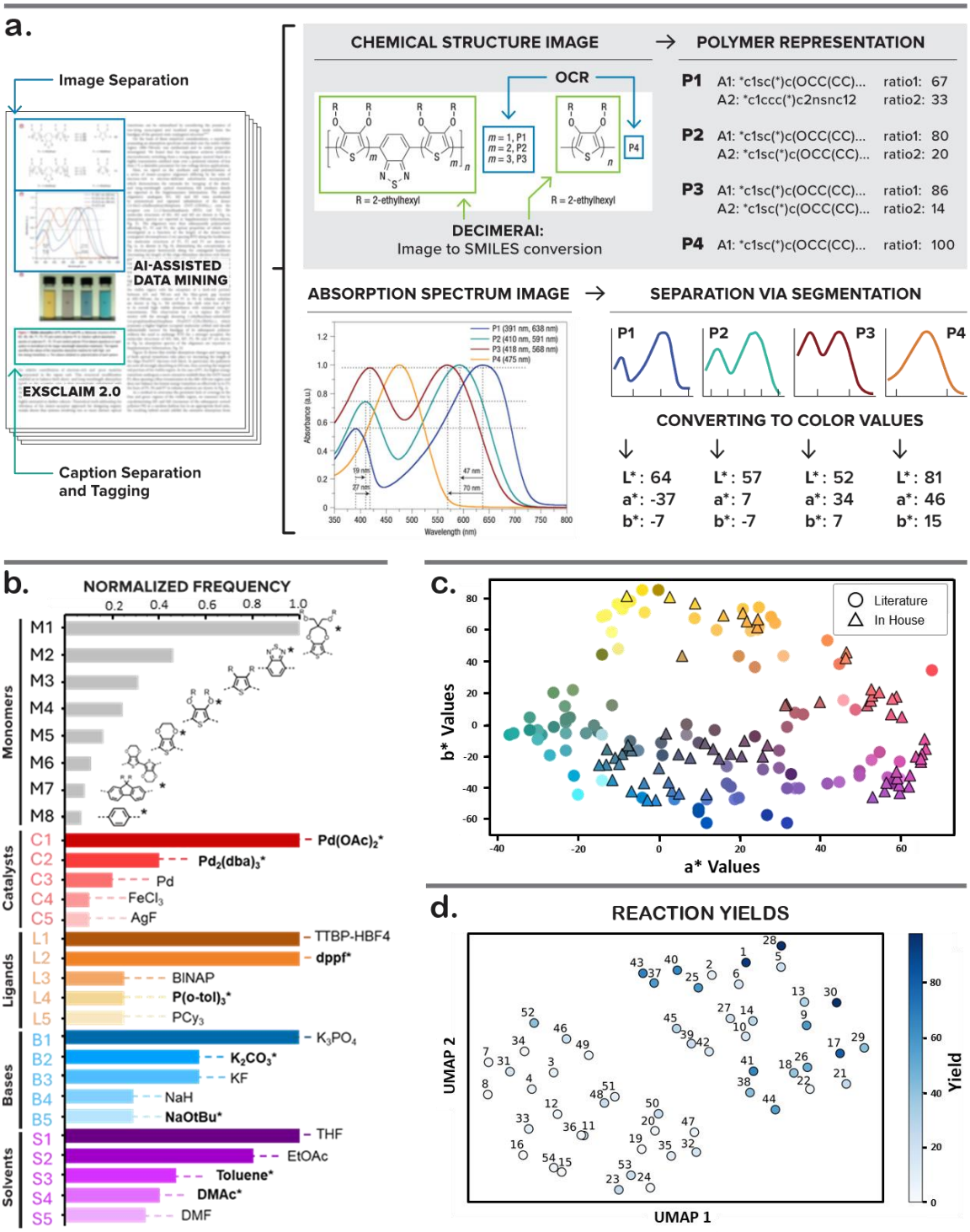

Fig. 2 Construction of the database and reaction optimization. (a) Example of the LLM-assisted data extraction process from ECP related literature; images in the original literature are first identified and categorized as 'absorption spectra' or 'molecular structure' by EXSCLAIM 2.0. The identified molecular structures are then converted to monomer SMILES strings using

DECIMER[43] and correlated with labels in the absorption spectra and ratio information using object character recognitions (OCR) methods. Finally, CIELab color (L*, a*, and b*) values are calculated from the digitalized absorption spectra. **(b)** Normalized frequencies of the most commonly used ECP monomers, catalysts, ligands, bases, and solvents in our literature searching database. Candidates used for later fine-tuning (M1, M2, M4, M5, M8) and autonomous discovery (M1, M2, M4, M5, M7 and M8) are marked by * symbols. **(c)** Scatterplot with the a*b* values of all the datapoints collected from literature (circle) and the in-house experimental batches (triangle). Color filled in represents the calculated color of polymer from absorption spectra. **(d)** Each of the categorical variables, base, solvent, ligand, catalyst is encoded and projected in 2-D using UMAP algorithm. The datapoints are color-coded based on the yield. The numbers above each datapoint indicate the corresponding reaction number in Table S2. The UMAP settings used are n_components=2, random_state=0, n_neighbors=10 and Euclidean distance as the distance metric.

Physics-informed predictive models for copolymer design

In the pursuit of predictive design of functional polymers, Polybot leverages a novel physics-informed model that incorporates two key innovations: a realistic copolymer representation and interpretable model outputs. Traditional sequence-based approaches typically encode only one instance at a time of the many possible monomer arrangements and therefore struggle with representing copolymers synthesized in experimental settings[44,45]. Since copolymer sequences cannot be fully controlled or characterized experimentally, even simple fixed-ratio linear copolymers exist as ensembles of energetically degenerate monomer arrangements. Our model addresses this challenge by employing a copolymer representation that encodes individual comonomer units along with their relative ratios, using the Set-Transformer architecture[46] – a self-attention autoencoder for sets (Fig. 3a, Supplementary Section 4.1-4.2 and Fig. S6). This model architecture ensures order-invariance in monomer inputs such that the combination of monomer A with monomer B is equivalent to the combination of monomer B with monomer A. Importantly, this approach can scale to more than two comonomer units without significantly increasing model

complexity. Additionally, our model is based on a dual-decoder design, which predicts both the color and absorption spectrum of ECPs. The color decoder outputs CIELab color values (L*, a*, and b*) with uncertainties obtained from Monte Carlo drop-out technique[47]. These outputs can be validated against experimental data or color values analytically calculated from the absorption spectrum predicted by the spectrum decoder. Simultaneously, the predicted spectrum provides insights into the relationship between monomer structures and their optical properties, enhancing model interpretability.

Our initial set of physics-informed models were trained exclusively on literature data from our ECP database (Table S1). During training, we employed a custom loss function that prioritizes the predictability of a* and b* (red/green/blue/yellow color) values over L* (perceptual lightness) values to focus on the perceived differences in hues. Among various combinations of input feature descriptors, molecular structural fingerprints such as the Morgan fingerprint[48] performs particularly well for encoding the model input, and these fingerprints perform even better in conjunction with electronic structure descriptions from density functional theory (DFT) calculations (Fig. 3a, Fig. S7, and Table S3).

To further enhance model performance, especially for ECPs with three types of comonomer units, we fine-tuned the model using three batches of in-house automated experiments (Fig. 3b, Supplementary Section 5), each consisting of 24 ECP samples. Our fine-tuning approach involved shortlisting five different types of comonomer units (Fig. 2b) based on their occurrence in the literature data, commercial availability, and synthesizability. Comonomer combinations were then selected from this list along with randomly assigned comonomer ratios. We structured the experimental batches as follows: the first batch with ECPs primarily copolymerized from two comonomer units, the second batch with both two and three comonomer units, and the third batch

with only three comonomer units (Tables S4-S6 and Fig. S8-S9). Several copolymer samples in different color themes were randomly selected to confirm their electrochromic properties, including onset potential of bleaching and optical contrast (Table S7 and Fig. S10-S11).

By incorporating high-throughput experimental results (Fig. 2c triangles), our fine-tuned physics-informed models outperform other common ML models, including typical artificial neural networks (ANN), random forest, XGBoost, Lasso, K-nearest neighbors (KNN), and gaussian process regression (GPR), in terms of both generalizability and color prediction accuracy (Fig. 3c). For a fair comparison, all models underwent fine-tuning with optimal hyperparameters using 5-fold cross validation (Fig. 3b, Tables S8-S9). To assess their performance, we split the combined literature and experimental batches dataset into an 80/20 training/test set, evaluating the mean absolute error (MAE) across the five folds (Fig. 3c). Remarkably, despite the small dataset used for training, our physics-informed model exhibits great precision after refinement (Fig. 3d). The color decoder model serves as our primary guide for autonomous experiments, owing to its low training data requirement and higher performance in color predictions, while the spectrum decoder model provides validation and extraction of insights from the predicted absorption peak information. We also created a user interface as a platform for a publicly accessible interactive ECP informatics and database, promoting collaboration and knowledge exchange within community[34] (Fig. 3e).

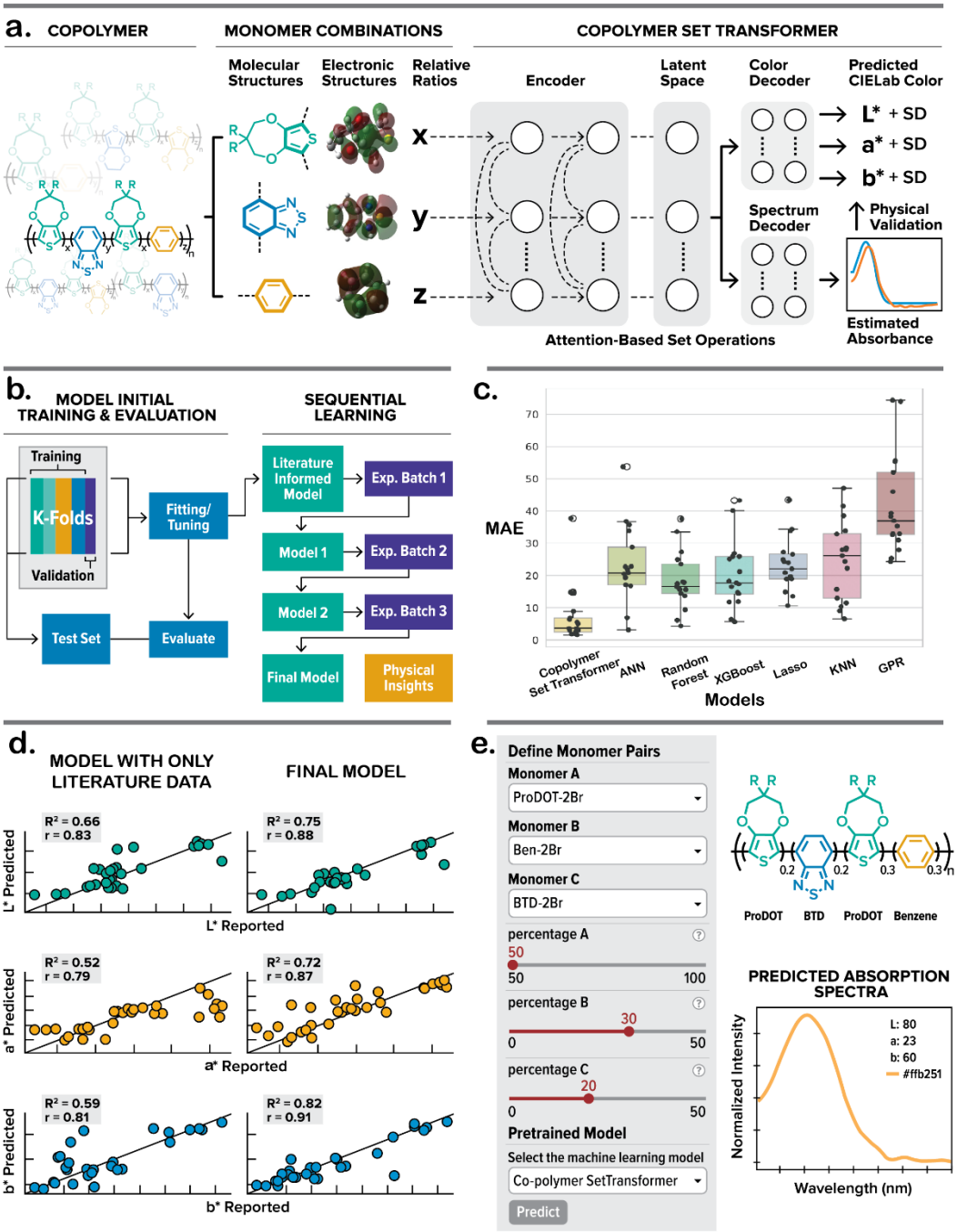

Fig. 3 The design of the physics-informed predictive model for ECPs. (a) Architecture of our physics-informed ML model. Input copolymer sequences are represented by monomer structure descriptors along with electronic structures and monomer ratios. Attention-based set operations

encode the monomer features into a latent space representation, which then simultaneously predicts CIELab color (L*, a*, and b*) color values and absorption spectra through a color decoder and spectrum decoder, respectively. **(b)** Workflow for hyperparameter tuning and validation of ML methods; a sequential learning approach enhances model performance using three batches of in-house experimental data. **(c)** Comparison of different multioutput regression models. Color prediction accuracy of these models is quantified using MAE on the hold-out test data. **(d)** The R^2 (coefficient of determination) value of our model, trained initially with literature only data (left) and fine-tuned with in-house experimental data (right). The parity plot shows the performance on the test data after splitting the train-test data with 80:20 ratio. **(e)** A publicly accessible user interface of our ECP predictive design toolkit[34].

Autonomous synthesis of copolymers with on-demand functionality

To autonomously produce functional polymers with high accuracy and efficiency, Polybot combines AI and robotic equipment in a continuous feedback loop, where experimental results are fed into our prediction model, allowing the AI to iteratively refine and optimize subsequent experiments in a self-driven manner (Fig. 4a). We set a goal to obtain ECPs with colors that are not in our database. We selected two target colors from less sampled regions of the color space. The color values are sent to the Polybot central computer, along with the available monomer units (labeled in Fig. 2b) and the desired percentage increment of their relative ratios (e.g., 5%) to initiate the autonomous synthesis of ECPs (Fig. S12). Polybot begins by querying the ECP database, using a previously trained ECP prediction model and the expected improvement policy to identify the six most promising formulas for the next iteration of experiments. Guided by the AI suggestions, Polybot then executes the robotic experimental workflow, seamlessly performing synthesis, purification, processing, material transfer, characterization, and data transfer (Supplementary Section 8 and Fig. S13-S14). The synthesis robot, UV-Vis characterization station, and transfer robot are coordinated within the generalizable network of the Polybot laboratory, which is built based on our modular science factory architecture[37] (Supplementary Movie 2 and 3). The

prediction model is iteratively retrained by using the newly updated database, with iterations continuing until the synthesized copolymers meet the color matching criteria we discuss below, and no further improvement is observed in the allocated budget (i.e., 24 copolymers based on the capacity of the robot module).

In just three iterations, Polybot efficiently explored the defined search space, autonomously synthesizing ECPs that exhibited closely matched target color values in both orange and green color themes (Fig. 4b). Color differences are quantified by using the delta-Lab (ΔE_{Lab}) metric, where smaller ΔE_{Lab} values indicate less perceptible color differences and values closer to 100 denote significant distortions. A ΔE_{Lab} value below 2.3 is barely discernible by the average human observer while a value between 3 and 6 is acceptable for most commercial reproduction[49,50]. This range is the target we set for this design. For the orange-theme ECPs targeting at (L* = 70, a* = 35, b* =50), Polybot produced all orange-colored ECPs in the first iteration and achieved ΔE_{Lab} values as small as 3.2 in subsequent iterations (Fig. 4b, Table S10 and Fig. S15, S17a). However, targeting for the green-colored ECPs (L* = 65, a* = -10, b* = 30), achieving the desired color value is more challenging due to the need for specific absorption profiles with similar intensity in the red (>600nm) and blue (<500nm) regions, and a sharp intensity decrease in the green region (500-600nm). These requirements, along with the limited green data points in our database, make the task difficult. However, despite initially sparse results, our algorithm converged in the final iteration and generated a new green ECP within our experimental design space with a ΔE_{Lab} of only 3.33 (Fig. 4b, Table S11 and Fig. S16, S17b). No further improvement was observed in iteration 3, so the Polybot terminated the autonomous search.

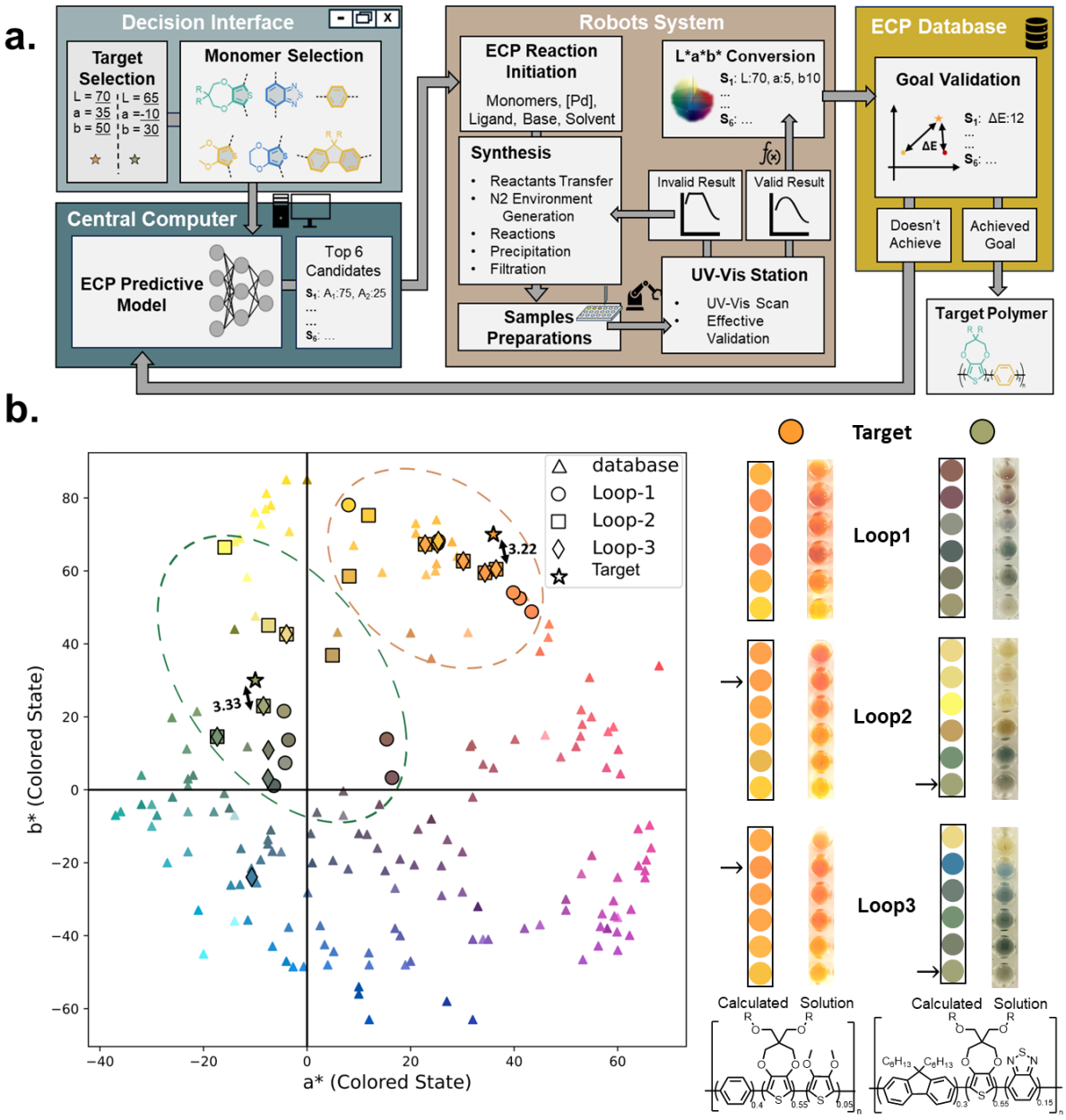

Fig. 4 Autonomous synthesis of ECPs with on-demand optical property with high accuracy and efficiency. (a) Workflow of our autonomous ECPs production involving autonomous execution of robotic experiments and decision logics. **(b)** Results from the autonomous production of ECPs targeting specific color values; Left: Color values of ECPs produced from each loop in the color space; the dashed circles group the data produced for each targeted color. Right: Visual comparison between the predicted colors and photos of ECP solutions from the three iteration loops of the autonomous discovery. The arrows indicate the optimal ECP results, and their chemical structures are shown at the bottom.

Conclusions

In summary, the accurate and precise synthesis of functional polymers has traditionally been labor-intensive and time-consuming. In this work, we introduced a strategy that boosts efficiency, thereby accelerating discovery and minimizing the need for repeated parameter adjustments and experimental trials. By leveraging LLM-assisted data mining, a novel copolymer Set-Transformer ML model, and a fully autonomous laboratory to substantially accelerate the on-demand synthesis of electronic polymers, we have achieved on-demand synthesis of the desired ECPs with great accuracy and efficiency. These AI-driven automated approaches advance the inverse design process through rapid knowledge accumulation, accurate property prediction, and fast experimental refinement using AI-guided robotic laboratories. Looking ahead, this platform's efficiency will be especially valuable for tackling underexplored targets, such as vivid greens and neutral greys, that demand unconventional polymer compositions. More broadly, this work highlights the potential of combining advanced AI with AI-driven automated laboratories to accelerate the design, synthesis, and time-to-market of diverse functional polymers with custom-engineered properties.

Methods

Data collection and database creation. ECP data were collected autonomously using the literature screening tool EXSCLAIM! and a modified version of the Plot2Spectra tool. The details are described in another publication[39,51]. Our literature screening is comprised of papers from Reynolds' group. This research group has extensive experience in ECP synthesis using simple synthetic steps and their papers are well-structured with absorption spectra images provided for all the polymers they report. Since it is not standard for materials researchers to release raw data along with their publications, we resorted to an interactive plot data extraction tool to extract data points from spectroscopy graph images. We did not use popular plot data extractor WebPlotDigitizer because that tool requires the user to align axes manually, input tick values, pick out the color of the target plot and draw the region in which the plot falls in a procedure that is difficult for large-scale data acquisition and that cannot be easily applied for automated data extraction[52]. Instead, we used Plot2Spectra to transform plot lines from a graph image into sets of coordinates in an automatic fashion. The plot digitizer comprises of two main steps, *i*) the axis alignment module to detect and recognize the ticks and find the region of interest and *ii*) the plot data extraction module to detect the plot lines. Filtering is based on anomaly detection. Overall, 100 ECPs were collected from 19 papers. A detailed list of these ECP pairs together with their characteristics and corresponding references are provided in the Supplementary Information (Table S1).

Color representation. For the creation of the electrochromic polymers database, we need to establish a reliable and consistent way to report the color and L*a*b* values given the ultraviolet-visible (UV-Vis) spectrum. Most of the L*a*b* values reported in literature refer to the colors of thin film results[31,33,36,53]. However, the absorption spectra refer to the solution state. For consistency in the L*a*b* values of our database we calculated all the L*a*b* values for the acquired absorption spectra using the Color python library with the following settings: MSDS_CMFS= CIE 2015 2 Degree Standard Observer, illuminant = D65.

Monomer representation. We explored different types of molecular representation for the monomers in our database, including molecular descriptors (2D and 3D), molecular fingerprint and DFT descriptors. Mordred 2D+3D software package which is freely available was employed to generate 1800 unique molecular descriptors for each compound directly from the SMILES strings[54]. For the extraction of the 3D molecular descriptors, we first generated the 3D

conformation of each monomer from the neutral SMILES and its geometry was optimized with the Merck Molecular Force Field (MMFF94) [55]. Mordred provides quick featurization of a molecular dataset by generating a vast array of two- and three-dimensional descriptor characteristics from the SMILES input. We trimmed down the 1800 descriptors to 1307 after removing those including NaN values. For the molecular fingerprint, the extended-connectivity fingerprint (ECPF) with length of 2048 bits and radius of three using the open-source RDKit package for Python was used[56]. ECFP provides a mechanism for representing topological chemical space by iteratively exploring substructure connectivity at a provided radius around each atom of a molecule. A detailed list of the 2D and 3D descriptors calculated with Mordred library is provided in the package documentation.[54]

DFT calculations. The Density Functional Theory (DFT) calculations were performed in an automated fashion employing the Auto-QChem workflow, which involves a) geometry optimization, b) frequency calculations and c) a time dependent DFT calculation for vertical excited state transitions[57]. Starting from the SMILES strings, the most stable conformer was generated using the RDKit package[56]. The hybrid B3LYP method and the 6-31G(d,p) basis set in gas phase was employed for all the calculations[58]. Time-dependent DFT (TD-DFT) was employed for the calculation of the excitation energies and optical properties. TD-DFT/B3LYP/6-31+G(d,p) was used for the ten first excited states as an efficient method for high-throughput screening of small organic molecules including C, O, N, S[59]. The input file with the calculation specifications and atomic coordinates for each of the individual monomers in our database was submitted to CNM's High Performance Computing system Carbon (https://wiki.anl.gov/cnm/HPC/), for DFT calculation using Gaussian 16.

Predictive AI. The built-in predictive model is based on the Set-Transformer architecture originally described in Lee et al[46]. We built, trained, and tested several architectures for each material representation. The input representation includes the N dimensional vectors that derive from the molecular representations as sets of instances also incorporating the ratio of each comonomer. To include the uncertainty estimation, we employed the Monte Carlo dropout technique to train with uncertainty without the need of training several Deep Neural Networks. The different models were Initially trained with the literature data and some experimental batches. The train objective was the MSE of the L*a*b* values. The network uses three neurons are the

output, one for each L*, a*, b* value. We also used a modified loss function such that the predictabilities of the a* and b* values are more important than the L* value.

As a distinctive feature of the architecture, each layer in the encoder and decoder attends to their inputs. Both the encoder and decoder are composed of stacks of neural network blocks with their own trainable parameters, i.e., attention-based set operations. Given a set of inputs, e.g., in our case the set of comonomers with their corresponding ratios, the attention module performs self-attention between the elements resulting in a set of equal size which contains information about the pairwise interactions among the elements in the input set. Then, several stacked attention-blocks in the encoder and decoder are able to capture the order invariance of the inputs and in the final layer the model outputs three values, L*, a*, and b* color coordinates.

The initial fine-tuning of the network was performed with k-fold cross validation on five randomly selected folds and the different models were compared based on the mean coefficient of determination (R^2) across the folds (Table S8). The selected hyperparameters for each model are described in detail in the (Table S9). The best performing model architecture with the tunned hyperparameters was selected and further re-trained on all the data and the saved model was used in the autonomous experiment to provide L*a*b* values prediction coupled with the uncertainty estimation. The model is initially trained on the literature data and keeps updating in an iterative manner after gathering the experimental results. After retraining, the ML model is able to provide a file with the suggested monomers and the ratios and the predicted color. We built a new architecture to deal with the input invariance problem using Set-Transformer and compared it with the widely used standard ML models. We tested the model by splitting the literature dataset in training/test set 80/20. We used the experimental data as an external validation set to measure the performance of entirely unknown ECPs and train different neural networks based on each representation. Our case is not the simple case of a single monomer or oligomer but involves combinations of comonomers and their corresponding ratios.

The two things we test are accuracy and generalizability. We tested various hyperparameters to select those that improve the model's generalizability, accuracy, learning rate and the ability for the network to learn the crucial features of the dataset. Although we started from a small and limited dataset, by retraining with high throughput experiments we managed to improve the generalizability and performance of our initial network and further use it for targeted exploration

on the experimental space. The loss function was modified so that the correct prediction of the a* and b* values is more significant than the prediction of the L* value.

Closed-loop experimentation. The robotic system worked with batches of 6 samples. It took 25 hours on average for the synthesis and purification procedure and 10 minutes to transfer and characterize the samples. We used a batched optimization strategy to select the 6 experiments in each round. The optimization algorithm is composed of Set-Transformer with prior knowledge as the surrogate model and expected improvement was the acquisition function taking as input the predicted deltaE (ΔE_{Lab}) of the targeted color and the uncertainty values (Supporting section 4, Fig S13-14). The whole process involves a full loop with synthesis and characterization and a data checking module after the loop for quality assessment. If the absorption spectra are not correct, we call a Chemspeed sub-process for diluting or adding extra sample to the sample positions.

User-friendly graphical user interface (GUI). The data were collected in an automated way and contain precise details of the experimental conditions, the absorption spectra measurements and the predicted and calculated L*a*b* values. The data were monitored and displayed live on a Streamlit dashboard. An outline of the dashboard and the GUI is presented in the Supporting Information (Fig. S12). The Streamlit app includes the literature database and provides a live monitoring of the experiments.

Robotic stations. For high-throughput experimentation and reliable replication in the context of automated new polymer discovery, a sophisticated array of robotic platforms was employed, each integrated with a specific role in the synthetic and analytical pipeline. Central to our automated synthesis operations was the Chemspeed Swing robot system (Chemspeed), a versatile workstation adept at precision dispensing of samples. Its strength, especially in polymer studies, is underscored by its proficiency in handling a wide spectrum of viscous solutions. Chemspeed's proprietary dosing technology and its capabilities of handling multiple reagent vials ensure both accuracy in delivery and reproducibility.

Following synthesis, the transfer of reaction mixtures for downstream processing was overseen by the UR5e Robotic Arm by Universal Robots (UR5e). Seamlessly integrated into our workflow, this robotic arm, with its six-axis mobility, offers an expansive range of motion, ensuring delicate polymer mixtures are maneuvered with precision and consistency. This ensures that samples experience minimal disturbance during transfers, an essential aspect that is often overlooked.

For analytical evaluations, we utilized the Tecan Infinite 200 Pro (Tecan) multimode plate reader. Known for its precision and versatility, the Tecan plate reader was equipped to deliver an array of detection methods, from fluorescence and absorbance to luminescence, rendering it an invaluable asset in the realm of polymer characterization. Beyond its detection abilities, its integration with advanced optics and automation software ensured not only high-resolution measurements but also high-throughput and consistent analysis.

In conclusion, the integrated use of these robotic platforms had optimized our system, enhancing our ability to uncover new polymers swiftly and methodically. Automation played a vital role, expanding our experimental capacity and ensuring consistent conditions—fundamental for data accuracy and reproducibility.

Equipment automation. Our experimental orchestration was streamlined through an integration of distinct robotic platforms—Chemspeed, UR5e, and Tecan—with the workflow execution interface (WEI), a product of Argonne National Laboratory[37]. WEI is a Python-based tool designed to enhance the automation and management of scientific tasks. It uses various executors, like the robotic operational system (ROS) and transmission control protocol (TCP) sockets, to improve communication between different modules[60]. One of WEI's key strengths is its ability to manage complex workflows in both research and lab environments. Workflows in WEI are outlined using the .yaml format. In this format, each workflow is a set of steps or commands, with each step carried out by a specific tool or component. This structured yet adaptable approach allows for the design and implementation of varied workflows across different research areas.

Both Tecan and Chemspeed were employed TCP sockets, optimized for synchronous communication, which continuously await robotic command directives dispatched from the WEI's TCP executor. Upon reception, these directives are processed and instantiated on the robots via their respective commercial software platforms. In practical terms, a WEI action directed towards these platforms might instruct the initiation of a specific protocol, accompanied by the requisite protocol file, or command the robot to commence reading results on a plate, delineating precise parameters for data acquisition. Protocol scripts were meticulously crafted utilizing the "pyautogui" library. These scripts, constructed by the synthetic chemists, employed "pyautogui" clicks and the execution of .exe files to dictate and sequence the requisite actions for each component. This

ensures real-time, low-latency execution of automation sequences, harmonizing WEI's capabilities with the robots' native software utilities.

In contrast, the UR5e platform is built around a ROS2 Node. This node actively listens for action messages dispatched by WEI and translates them into robot actions using our UR driver package. This package, a combination of specialized sub-modules, is engineered to provide granular control over the robot's capabilities. Notably, it interfaces with the URx Python library, a middleware that abstracts and executes intricate operations, from Cartesian transfers to pick-and-place sequences. Complementing this, the UR driver also integrates with the UR dashboard interface, enabling advanced functionalities encompassing robot initialization and execution of the ".urp" programs. The control mechanism for the UR5e arm was delineated through the creation of two ".urp" files. These files contain the specific movement sets essential for transporting the plate between the Chemspeed and Tecan units, ensuring seamless bidirectional mobility.

Users architect an experimental workflow outlining each step, articulated in a detailed format. Subsequently, WEI executors parse and instantiate these directives, invoking the required executor—either TCP or ROS—tailored for the specific robotic platform's communication protocol.

Process management system. Within our process management system, three distinct workstations are prominently featured, each utilized for a specific operational function. One workstation is dedicated to the meticulous control of Chemspeed, ensuring precision in its operations. A second is tailored for the management of Tecan, optimizing its functionalities. The third workstation holds a key position: it orchestrates the entire workflow by executing the WEI workflows, effectively serving as the central command hub for the entire orchestration process. Here, workflow steps are diligently parsed, and subsequent action calls are intelligently dispatched to the respective instruments. Notably, the UR robot, seamlessly integrated into our lab's network infrastructure, can be directly communicated from this central workstation, ensuring real-time coordination.

The Globus Architecture for Data-Intensive Experimental Research (Gladier) played an important role in our data management and analysis framework. Gladier was elaborately employed to design a dataflow, facilitating a seamless integration of the multiple data sources present within our laboratory environment. Upon the completion of readings by the Tecan plate reader, the generated

data is transferred to the Globus database via its dedicated Globus endpoint. However, the data's progression doesn't stop at that point. The master computer, through its dedicated Globus endpoint, then fetches this information. With the data securely in its precincts, it's primed for the subsequent execution of machine learning algorithms, ensuring a holistic and streamlined analytical process. Once safely ensconced within the Globus database, our machine learning algorithms are invoked, leveraging the capabilities of Globus Flows[61]. This platform is adept at handling sophisticated algorithms, enabling us to derive insights and patterns from the captured data with heightened accuracy and reliability. This integration highlights our dedication to a unified strategy in data acquisition, safe storage, and thorough analysis.

Thin film and device fabrication. For the single-colored thin film, ECP films were spin-coated on cleaned indium tin oxide (ITO) substrates with ECP solution 20 mg/mL dissolved in chloroform at 1500 rpm. For the flexible ECPs display, ECPs were first dissolved separately in 9:1 of chlorobenzene: N-methyl-2-pyrrolidone solution separately at 4mg/mL to prepare the ECP inks. These polymer inks were printed into the designed patterns on the polyethylene terephthalate (PET)-ITO substrates using an Optomec Aerosol Jet 5X system. A gel electrolyte was then prepared by mixing poly(ethylene glycol) methacrylate (Mn = 500), 0.2 M solution of 1-ethyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide in propylene carbonate, and 2-hydroxy-2-methylpropiophenone in a volume ratio of 5:5:1, respectively. This gel electrolyte was drop-casted onto a bare PET-ITO substrate, serving as anode, followed by placing the PET-ITO with printed ECP patterns on top of the electrolyte. The device was irradiated under a UV lamp for 10 min to crosslink the gel electrolyte. Finally, the display was operated under 2 V.

References:

1 Banerjee, T., Podjaski, F., Kröger, J., Biswal, B. P. & Lotsch, B. V. Polymer photocatalysts for solar-to-chemical energy conversion. *Nat. Rev. Mater.* **6**, 168-190 (2021).

2 Fratini, S., Nikolka, M., Salleo, A., Schweicher, G. & Sirringhaus, H. Charge transport in high-mobility conjugated polymers and molecular semiconductors. *Nat. Mater.* **19**, 491-502 (2020).

3 Qiu, Z., Hammer, B. A., & Müllen, K.. Conjugated polymers–Problems and promises. Progress in Polymer *Science*, **100**, 101179 (2020).

4 Szymanski, N. J. *et al.* An autonomous laboratory for the accelerated synthesis of novel materials. *Nature* **624**, 86-91 (2023).

5 Ha, T. *et al.* AI-driven robotic chemist for autonomous synthesis of organic molecules. *Sci. Adv.* **9**, eadj0461 (2023).

6 Koscher, B. A. *et al.* Autonomous, multiproperty-driven molecular discovery: From predictions to measurements and back. *Science* **382**, eadi1407 (2023).

7 Burger, B. *et al.* A mobile robotic chemist. *Nature* **583**, 237-241 (2020).

8 Coley, C. W. *et al.* A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* **365**, eaax1566 (2019).

9 Granda, J. M., Donina, L., Dragone, V., Long, D.-L. & Cronin, L. Controlling an organic synthesis robot with machine learning to search for new reactivity. *Nature* **559**, 377-381 (2018).

10 Strieth-Kalthoff, F. *et al.* Delocalized, asynchronous, closed-loop discovery of organic laser emitters. *Science* **384**, eadk9227 (2024).

11 Montoya, J. H. *et al.* Toward autonomous materials research: Recent progress and future challenges. *Appl. Phys. Rev.* **9** (2022).

12 Angello, N. *et al.* Closed-loop transfer enables ai to yield chemical knowledge. Preprint at https://chemrxiv.org/engage/chemrxiv/article-details/64ef56463fdae147fa2346d4 (2023).

13 Lyu, H., Ji, Z., Wuttke, S. & Yaghi, O. M. Digital reticular chemistry. *Chem* **6**, 2219-2241 (2020).

14 Yang, Q. *et al.* Artificial intelligence for conjugated polymers. *Chem. Mater.* **36**, 2602-2622 (2024).

15 Jafari, V. F., Mossayebi, Z., Allison-Logan, S., Shabani, S. & Qiao, G. G. The power of automation in polymer chemistry: Precision synthesis of multiblock copolymers with block sequence control. *Chem. – Eur. J.* **29**, e202301767 (2023).

16 Patel, R. A. & Webb, M. A. Data-driven design of polymer-based biomaterials: High-throughput simulation, experimentation, and machine learning. *ACS Appl. Bio Mater.* **7**, 510-527 (2024).

17 Martin, T. B. & Audus, D. J. Emerging trends in machine learning: A polymer perspective. *ACS Polym. Au* **3**, 239-258 (2023).

18 Yan, C. & Li, G. The rise of machine learning in polymer discovery. *Adv. Intell. Syst.* **5**, 2200243 (2023).

19 Vriza, A., Chan, H. & Xu, J. Self-driving laboratory for polymer electronics. *Chem. Mater.* **35**, 3046-3056 (2023).

20 Lo, S., Seifrid, M., Gaudin, T. & Aspuru-Guzik, A. Augmenting polymer datasets by iterative rearrangement. *J. Chem. Inf. Model.* **63**, 4266-4276 (2023).

21 Audus, D. J. & de Pablo, J. J. Polymer informatics: Opportunities and challenges. *ACS Macro Lett.* **6**, 1078-1082 (2017).

22 Callaway, C. P. *et al.* The solution is the solution: Data-driven elucidation of solution-to-device feature transfer for π-conjugated polymer semiconductors. *ACS Appl. Mater. Interfaces* **14**, 3613-3620 (2022).

23 Lee, J. *et al.* A fully automated platform for photoinitiated raft polymerization. *Digital Discovery* **2**, 219-233 (2023).

24 Schuett, T., Kimmig, J., Zechel, S. & Schubert, U. S. Automated polymer purification using dialysis. *Polymers* **12**, 2095 (2020).

25 Reis, M. *et al.* Machine-learning-guided discovery of 19F MRI agents enabled by automated copolymer synthesis. *J. Am. Chem. Soc.* **143**, 17677-17689 (2021).

26 Knox, S. T., Parkinson, S. J., Wilding, C. Y. P., Bourne, R. A. & Warren, N. J. Autonomous polymer synthesis delivered by multi-objective closed-loop optimisation. *Polym. Chem.* **13**, 1576-1585 (2022).

27 MacLeod, B. P. *et al.* Self-driving laboratory for accelerated discovery of thin-film materials. *Sci. Adv.* **6**, eaaz8867 (2020).

28 Rubens, M., Vrijsen, J. H., Laun, J. & Junkers, T. Precise polymer synthesis by autonomous self-optimizing flow reactors. *Angew. Chem., Int. Ed.* **58**, 3183-3187 (2019).

29 Vieira, R. A. M., Sayer, C., Lima, E. L. & Pinto, J. C. Closed-loop composition and molecular weight control of a copolymer latex using near-infrared spectroscopy. *Ind. Eng. Chem. Res.* **41**, 2915-2930 (2002).

30 Tom, G. *et al.* Self-driving laboratories for chemistry and materials science. Preprint at https://chemrxiv.org/engage/chemrxiv/article-details/66708e2e5101a2ffa8cbd354 (2024).

31 Beaujuge, P. M., Ellinger, S. & Reynolds, J. R. The donor–acceptor approach allows a black-to-transmissive switching polymeric electrochrome. *Nat. Mater.* **7**, 795-799 (2008).

32 Li, X., Perera, K., He, J., Gumyusenge, A. & Mei, J. Solution-processable electrochromic materials and devices: Roadblocks and strategies towards large-scale applications. *J. Mater. Chem. C* **7**, 12761-12789 (2019).

33 You, L., He, J. & Mei, J. Tunable green electrochromic polymers via direct arylation polymerization. *Polym. Chem.* **9**, 5262-5267 (2018).

34 Vriza, A. Polybot database: Electrochromic polymers. https://polybot-ecps.streamlit.app/ (2024).

35 Beaujuge, P. M. & Reynolds, J. R. Color control in π-conjugated organic polymers for use in electrochromic devices. *Chem. Rev.* **110**, 268-320 (2010).

36 Christiansen, D. T., Ohtani, S., Chujo, Y., Tomlinson, A. L. & Reynolds, J. R. All donor electrochromic polymers tunable across the visible spectrum via random copolymerization. *Chem. Mater.* **31**, 6841-6849 (2019).

37 Vescovi, R. *et al.* Towards a modular architecture for science factories. *Digital Discovery* **2**, 1980-1998 (2023).

38 Tabak, T., Altinisik, S., Ulucay, S., Koyuncu, S. & Kaya, K. Black-to-transmissive electrochromic switching pedot-co-poly(n-ethylcarbazole) via a sustainable and facile in situ photo(co)polymerization method. *Macromolecules* **57**, 4769-4781 (2024).

39 Schwenker, E. *et al.* Exsclaim!: Harnessing materials science literature for self-labeled microscopy datasets. *Patterns* **4**, 100843 (2023).

40 Österholm, A. M., Shen, D. E. & Reynolds, J. R. Electrochromism in conjugated polymers – Strategies for complete and straightforward color control in *Conjugated polymers* 201-248 (CRC Press, 2019).

41 Schanda, J. CIE Colorimetry in *Colorimetry: Understanding the CIE system* Ch. 3, (John Wiley & Sons, 2007).

42 Chen, X., Engle, K. M., Wang, D.-H. & Yu, J.-Q. Palladium(II)-catalyzed C-H activation/C-C cross-coupling reactions: Versatility and practicality. *Angew. Chem., Int. Ed.* **48**, 5094-5115 (2009).

43 Rajan, K., Brinkhaus, H. O., Agea, M. I., Zielesny, A. & Steinbeck, C. Decimer.Ai: An open platform for automated optical chemical structure identification, segmentation and recognition in scientific publications. *Nat. Commun.* **14**, 5045 (2023).

44 Patel, R. A., Borca, C. H. & Webb, M. A. Featurization strategies for polymer sequence or composition design by machine learning. *Mol. Syst. Des. Eng.* **7**, 661-676 (2022).

45 Lin, T.-S. *et al.* Bigsmiles: A structurally-based line notation for describing macromolecules. *ACS Cent. Sci.* **5**, 1523-1531 (2019).

46 Lee, J. *et al.* Set transformer: A framework for attention-based permutation-invariant neural networks in *International conference on machine learning*. 3744-3753 (2019).

47 Hirschfeld, L., Swanson, K., Yang, K., Barzilay, R. & Coley, C. W. Uncertainty quantification using neural networks for molecular property prediction. *J. Chem. Inf. Model.* **60**, 3770-3780 (2020).

48 Gardiner, E. J., Holliday, J. D., O'Dowd, C. & Willett, P. Effectiveness of 2D fingerprints for scaffold hopping. *Future Med. Chem.* **3**, 405-414 (2011).

49 Sharma, G., Wu, W. & Dalal, E. N. The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Res. Appl.* **30**, 21-30 (2005).

50 Chung, R., Yu, L. & Myers, B. The effect of OBA in paper and illumination level on perceptibility of printed colors in *TAGA Annual Technical Conference* (2015).

51 Jiang, W. *et al.* Plot2spectra: An automatic spectra extraction tool. *Digital Discovery* **1**, 719-731 (2022).

52 Marin, F., Rohatgi, A. & Charlot, S. Webplotdigitizer, a polyvalent and free software to extract spectra from old astronomical publications: Application to ultraviolet spectropolarimetry. Preprint at https://arxiv.org/abs/1708.02025 (2017).

53 Beaujuge, P. M., Ellinger, S. & Reynolds, J. R. Spray processable green to highly transmissive electrochromics via chemically polymerizable donor-acceptor heterocyclic pentamers. *Adv. Mater.* **20**, 2772-2776 (2008).

54 Moriwaki, H., Tian, Y.-S., Kawashita, N. & Takagi, T. Mordred: A molecular descriptor calculator. *J. Cheminf.* **10**, 4 (2018).

55 Halgren, T. A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of mmff94. *J. Comput. Chem.* **17**, 490-519 (1996).

56 Landrum, G. Rdkit: A software suite for cheminformatics, computational chemistry, and predictive modeling. http://www.rdkit.org/RDKit_Overview.pdf (2013).

57 Żurański, A. M., Wang, J. Y., Shields, B. J. & Doyle, A. G. Auto-QChem: An automated workflow for the generation and storage of DFT calculations for organic molecules. *React. Chem. Eng.* **7**, 1276-1284 (2022).

58 Becke, A. D. A new mixing of Hartree–Fock and local density-functional theories. *J. Chem. Phys.* **98**, 1372-1377 (1993).

59 Tirado-Rives, J. & Jorgensen, W. L. Performance of B3LYP density functional methods for a large set of organic molecules. *J. Chem. Theory Comput.* **4**, 297-306 (2008).

60 Quigley, M. *et al.* ROS: an open-source Robot Operating System in *ICRA workshop on open source software.* 5 (Kobe, Japan).

61 Chard, R. *et al.* Globus automation services: Research process automation across the space–time continuum. *Future Gener. Comput. Syst.* **142**, 393-409 (2023).

Acknowledgments: Work performed at the Center for Nanoscale Materials, a U.S. Department of Energy Office of Science User Facility, was supported by the U.S. Department of Energy (DOE), Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357, and Laboratory Directed Research and Development (LDRD) funding from Argonne National Laboratory, provided by the Director, Office of Science, of the U.S. DOE under Contract No. DE-AC02-06CH11357. J.X. acknowledges the Big Ideas Generator (BIG) seed funding program from the University of Chicago. A.M.Ö and J.R.R. acknowledge the Air Force Office of Scientific Research for financial support (FA9550-25-1-0021). The authors thank Center for Nanoscale Materials, a U.S. Department of Energy Office of Science User Facility, for performing this work. The authors thank the IT team at the Center for Nanoscale Materials, particularly R. McQueen, and M. Welch, for their assistance with the lab network. The authors also thank the Materials Engineering

Research Facility at Argonne National Lab for providing electronic printing support. The authors thank H. C. Fry for helping to integrate Tecan spectrophotometer. The authors thank the contributions of the following students and postdocs: X. Li, J. Fan, C. Castro, D, Simatos, F. Kim, A. Koneru, S. Banik, B. J. Chaudhary, J. Zhou, M. Palak, S. Lee, Y. Li, R. Papani, N. Shan, M. Weires, J. Li, Z. Liu, C. Zhang, W. Liu, R. Yetishefsky from the University of Chicago, Purdue University, University of Illinois at Chicago, and Argonne National Laboratory for testing the accuracy of the data mining toolkit.

Author contributions:

J.X., Y.W., and A.V. conceived the research. Y.W., J.Z., Q.Y., and J.M. designed and performed the synthesis work. A.V., M.K.Y.C., and H.C. performed the literature data mining and constructed the ML models. D.O., A.V., R.V., and I. F. set up the SDL network. Z.W. and Y.W. carried electrochromic characterizations. S.H. and Y.Z. helped with printing demonstrations. A.M.Ö. and J.R.R. contributed to the database construction. S.K.R. S.S. assisted with the initial conceptualization of Polybot. J.X., H.C., and J.M., supervised the research. Y.W., A.V., H.C., J.M., and J.X. wrote the manuscript. All the authors contributed to the discussion and manuscript revision.

Data availability:

The data supporting this study are available in the main text, Supplementary Information, and GitHub repository at https://github.com/polybot-nexus/EPoly_LitMining_ColorPred/tree/main/datasets. The electrochromic polymer datasets used in our analyses can be found in Supplementary Information sections 2 and 5 while all synthesis

optimization data are provided in Supplementary Information sections section 3. Detailed experimental protocols for data extraction and robotic laboratory operations are also included in the Supplementary Information.

Code availability:

Codes and demonstrations related to the LLM-assisted text/image data mining workflows, model training, and descriptor calculations are available on GitHub at https://github.com/polybot-nexus/EPoly_LitMining_ColorPred, along with a detailed README of the required Python packages and installation instructions. An interactive web app is accessible at https://polybot-ecps.streamlit.app/ to demonstrate ECP color and absorption spectra prediction based on the Co-polymer Set Transformer model.

Additional Information

Supplementary Information is available for this paper.

Competing interests: The authors declare no competing interests.

Correspondence and requests for materials should be addressed to J.X., J.M., and H. C. (xuj@anl.gov; jgmei@purdue.edu; hchan@anl.gov)

.