ISSN: 0976-7126
Chaturvedi & Gupta, 16(12):15-19, 2025

International Journal of Pharmacy & Life Sciences
Open Access to Researcher

AI-Driven Discovery of Novel Polymers: A Comprehensive Review

Prerna Chaturvedi*and Arun Kumar Gupta

Chameli Devi Institute of Pharmacy, Indore, (M.P) - India

Article info

Received: 18/10/2025

Revised:20/11/2025

Accepted: 19/12/2025

© IJPLS

www.ijplsjournal.com

Abstract

The discovery and optimization of polymeric materials have traditionally relied on experiment-driven, trial-and-error workflows that are time-consuming and resource intensive. Artificial intelligence (AI) — particularly machine learning (ML), graph neural networks (GNNs), generative models, and active learning — is transforming polymer science by enabling rapid property prediction, inverse design, and closed-loop experimental optimization. This review surveys recent advances (2018–2025) in data resources, polymer representations, predictive and generative AI models, optimization strategies (including Bayesian optimization and active learning), and automated/self-driving laboratories. We highlight landmark platforms (Polymer Genome, Open Macromolecular Genome), methodological progress (multitask GNNs, chemical language models such as polyBERT, and benchmarks for deep generative models), and successful demonstrations of AI-assisted polymer discovery in energy, electronics, healthcare, and sustainable plastics.

Key challenges remain: scarcity and heterogeneity of polymer data, difficulty representing polymer ensembles and architectures, synthetic feasibility of generated candidates, model interpretability, and integration with experimental workflows. We discuss strategies to address these issues — standardized databases and representations, self-supervised learning, synthesis-aware generative models, and tighter AI–robotics integration — and outline opportunities where AI can accelerate green polymer chemistry, circular-economy polymers, and application-directed multi-property optimization. The review concludes by advocating for community efforts in data curation, open benchmarks, and interdisciplinary training to realize AI's promise for fast, cost-effective polymer innovation.

Keywords: Polymer, AI, Material

Introduction

Polymers underpin an enormous fraction of modern technologies — from packaging and structural materials to biomedical devices and organic electronics. Their property space is vast because small changes in monomer chemistry, sequence, tacticity, and processing produce large changes in performance. Traditional discovery cycles (synthesis → characterization → iterative design) are slow and expensive. Over the past decade AI has emerged as a powerful complement to experiment and simulation, enabling (i) rapid property prediction from structure, (ii) inverse design to identify candidate chemistries that meet target properties, and (iii) closed-loop optimization when paired with automated synthesis and characterization platforms.

***Corresponding Author**
E.mail: prernachaturvedi12@gmail.com

ISSN: 0976-7126
Chaturvedi & Gupta, 16(12):15-19, 2025

Landmark initiatives — notably the Polymer Genome platform — established the feasibility of ML-driven property prediction for polymers and motivated subsequent methodological innovations (e.g., graph neural networks and chemical language models) that scale polymer informatics to much larger candidate spaces. [1]

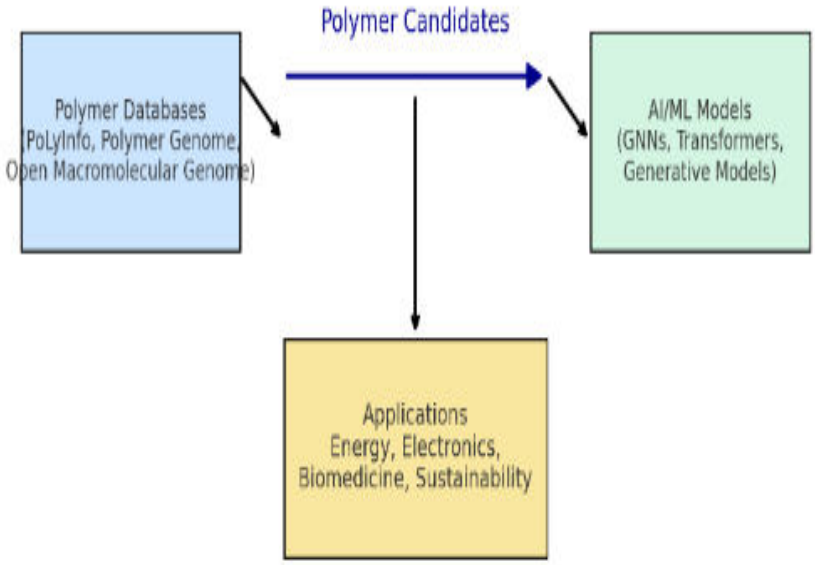

Fig. 1: Block diagram of AI driven Novel Polymer

Data resources and curation

AI requires curated data. For polymers, important sources include PoLyInfo (NIMS), Polymer Genome datasets, the Open Macromolecular Genome (OMG), CAMPUS and commercial datasheets, and public repositories such as PubChem-derived polymer entries. PoLyInfo aggregates >0.5M polymer data points with measurement metadata; Polymer Genome supplies curated property labels and models for many thermomechanical properties; OMG focuses on synthetic accessibility and reaction-compatibility for generated chemistries. Data challenges are numerous: disparate measurement conditions, inconsistent naming/representation, and limited coverage of copolymers and complex architectures. Recent reviews and database efforts emphasize machine-readable standards and the need for richer metadata (processing, degree of polymerization, measurement conditions) to improve model transferability. [2-3]

Representations and descriptors for polymers

A core difficulty in polymer informatics is representing polymers compactly yet expressively for ML. Early approaches used handcrafted features derived from monomer fingerprints, constitutional descriptors, and thermodynamic approximations (the Polymer Genome fingerprint). More recent representations learn features directly using graph neural networks (GNNs) and sequence/transformer-based encodings (chemical language models). Graph-based representations that capture monomer graphs, connection topology, and periodicity have shown improved accuracy and enabled multitask learning across many properties. Self-supervised pretraining (e.g., polyBERT-style models) further improves data efficiency and enables transfer learning across tasks. Still, representing copolymer sequence distributions, polydispersity, and chain architecture remains an active research area. [4]

Predictive ML models: property forecasting and screening

Predictive models (regression/classification) remain the first AI tool in polymer discovery. Approaches include random forests, gradient-boosted trees, fully connected neural networks,

and GNNs. Multitask learning — training one model to predict many properties — leverages shared structure–property correlations and reduces required labeled data. GNN-based models have been shown to speed feature extraction and screening by 1–2 orders of magnitude while maintaining or improving accuracy relative to handcrafted descriptors, enabling virtual screening of millions of candidates. These surrogate models are widely used as fast filters upstream of synthesis planning.

Generative models and inverse design

Inverse design seeks to produce polymer chemistries that match desired properties. Generative models used include Variational Autoencoders (VAEs), Generative Adversarial Networks (GANs), recurrent neural networks, reinforcement learning (RL) strategies, and graph-based generative models. Recent efforts emphasize *synthesis awareness* (ensuring generated chemistries are compatible with known polymerization reactions and available monomers) — the Open Macromolecular Genome (OMG) is an explicit example that constrains generation to synthetically accessible chemistries. Benchmarking studies of deep generative models for polymers have begun to evaluate validity, novelty, diversity, and property-guided optimization performance, highlighting tradeoffs between exploration and synthetic realism. [5-6]

Optimization, active learning, and Bayesian methods

When experimental throughput is limited, optimization strategies such as Bayesian optimization (BO) and active learning (AL) are extremely valuable. BO has been applied to tune polymerization conditions and compositions for target mechanical or thermal outcomes, often with acquisition functions tailored to multi-objective goals. Batch and scalable BO variants permit efficient exploration of high-dimensional polymer formulation spaces. Coupling BO with predictive surrogates and uncertainty quantification accelerates discovery while minimizing costly experiments.

Integration with automation: self-driving laboratories

The ultimate acceleration arises when AI models are integrated with automated synthesis and characterization — so-called self-driving laboratories (SDLs). SDLs close the loop: the model proposes candidates, robotics synthesize them, instruments measure properties, and the data are fed back to update the model. Recent high-impact demonstrations in materials science show autonomous laboratories optimizing inorganic and soft materials properties orders of magnitude faster than manual workflows; similar platforms have been developed specifically for polymers (e.g., Polybot and other autonomous polymer labs), enabling rapid optimization of electronic and soft-material properties. SDLs also demand careful experimental design, standardized protocols, and tight error-handling to ensure robustness. [7-8]

Selected application areas and success stories

Energy and electronics — AI has accelerated discovery of high-performance polymer dielectrics, polymer electrolytes for batteries, and conjugated polymers for organic electronics through combined modeling and high-throughput screening.

Sustainable and biodegradable polymers — Active learning and BO have guided formulation of biodegradable polyesters with improved degradation vs. mechanical property tradeoffs.

Biomedical polymers — AI-guided design of hydrogels and drug-delivery polymers helps tailor swelling, release kinetics, and biocompatibility while reducing animal testing. (Emerging; several proof-of-concept demonstrations exist.)

Catalyst and process optimization — AI has assisted in selecting polymerization catalysts (stereoselective ROP) and optimizing process conditions for desired tacticity and molecular weight distribution. [9]

Challenges and limitations

Data quality and quantity: Polymer datasets are noisy, sparse, and heterogeneous. Measurement conditions (e.g., molecular weight, processing history) are often missing or inconsistent, limiting generalizability.

Representation of polymer ensembles: Unlike small molecules, polymers are distributions with polydispersity and sequence variability. Representations that fully capture these features are still maturing.

Synthetic feasibility: Generative models can produce chemically valid but synthetically infeasible polymers. Constraining generation to reaction-compatible chemistries or embedding synthesis planning is necessary.

Interpretability and trust: Many deep models are black boxes; explainable AI and physically informed ML are required for adoption by chemists and engineers.

Integration costs: Building and operating SDLs is nontrivial, requiring robotics, instrumentation, and software orchestration. There are also safety and reproducibility considerations. [10-11]

Paths forward and recommendations

Standards and community datasets: Community efforts should standardize polymer metadata, measurement reporting, and open benchmark datasets to enable fair comparisons. PoLyInfo, Polymer Genome, and OMG are positive examples.

Representation research: Continue developing representations that capture sequence, architecture, and ensemble effects (periodic graphs, augmented GNNs, and transformer encodings).

Synthesis-aware generative models: Embed reaction templates, monomer availability, and polymerization mechanism constraints in generative workflows to prioritize accessible candidates.

Self-supervised and transfer learning: Use large unlabeled polymer corpora for pretraining (polyBERT, self-supervised GNNs) to improve data efficiency when labeled data are scarce.

Interdisciplinary training & open tooling: Broaden training across polymer chemistry, AI, and automation; produce open toolchains and reproducible workflows (model code, datasets, and SDL software). [12-14]

Conclusion

AI has matured from speculative promise to a practical accelerator of polymer discovery. Advances in database construction, polymer-specific representations, multitask GNNs, chemical language models, synthesis-aware generative frameworks, and autonomous laboratories together enable faster, cheaper, and more directed polymer innovation. To fully realize this promise, the community must tackle data standardization, representation of polymer complexity, synthetic realism, explainability, and practical integration into experimental workflows. The coming decade should witness broader adoption of AI-guided pipelines that deliver application-ready polymers for energy, healthcare, electronics, and a circular plastics economy.

References

1. Kim C, Chandrasekaran A, Huan TD, Das D, Ramprasad R. Polymer Genome: A data-powered polymer informatics platform for property predictions. *J Phys Chem C.* 2018;122(31):17575–17585.
2. Gurnani R, Kuenneth C, Toland A, Ramprasad R. Polymer informatics at scale with multitask graph neural networks. *Chem Mater.* 2023;35(4):1560–1567.
3. Kuenneth C, Chandak R, Ramprasad R. polyBERT: a chemical language model to enable fully data-driven polymer informatics. *Nat Commun.* 2023;14.
4. Kim S, et al. Generative design of synthetically accessible polymers. *ACS Polymers Au.* 2023.
5. Tom G, et al. Self-driving laboratories for chemistry and materials science. *Chem Rev.* 2024.
6. Szymanski NJ, et al. An autonomous laboratory for the accelerated synthesis of inorganic materials. *Nature.* 2023.
7. Wang X, et al. Bayesian-optimization-assisted discovery of catalysts/processes (example application in materials). *Nat Commun.* 2023.

8. Ishii M. NIMS PoLyInfo polymer database review: PoLyInfo (I & II). *J Mater Inf.* 2024.
9. Reiser P, et al. Graph neural networks for materials science and chemistry. *Comput Mater.* 2022;
10. Yue T, Tao L, Varshney V, Li Y. Benchmarking study of deep generative models for inverse polymer design. *Digital Discovery / ChemRxiv.* 2024;
11. Low AKY, et al. Self-driving laboratories: Translating materials science into autonomous discovery. *Nat Rev Mater.* 2025.
12. Pilania G, et al. Representative early works on Polymer Genome & property prediction methods). *J Appl Phys.* 2020; Queen O, et al. Polymer graph neural networks for multitask property learning. *npj Comput Mater.* 2023.
13. Jansen SAH, et al. Bayesian optimization for multicomponent supramolecular/materials design (methodological example). *J Am Chem Soc.* 2019.
14. Kassab A, et al. Graph neural network-driven surrogate modeling for accelerated polymer modeling (recent advances). *Comput Mater Sci.* 2025.

Cite this article as:
Chaturvedi P. and Gupta A. K. (2025). AI-Driven Discovery of Novel Polymers: A Comprehensive Review. *Int. J. of Pharm. & Life Sci.,* 16(12):15-19.

Source of Support: Nil
Conflict of Interest: Not declared
For reprints contact: ijplsjournal@gmail.com