

AUTONOMOUS POLYMER MATERIALS DISCOVERY: ARTIFICIAL INTELLIGENCE ARCHITECTURES FOR GENERATIVE DESIGN AND CLOSED-LOOP ENGINEERING

Victor Ugabetan AGBOGO^{1,*}[0009-0002-3028-2005], Moipone Linda TEFFO¹[0000-0003-3000-1269], Lucey Mapula MAVHUNGU¹[0000-0003-3000-1269], Emmanuel Rotimi SADIKU¹[0000-0002-8504-1041],
Orinya M. INJOR¹[0009-0008-9453-0948]

¹Department of Chemical, Metallurgical, and Materials Engineering (Polymer Division) & Institute of Nano-Engineering Research (INER), Tshwane University of Technology, Pretoria 0001, South Africa

Abstract

Artificial intelligence (AI) is revolutionizing polymer engineering by shifting materials development from traditional trial-and-error to integrated, data-driven, and autonomous workflows. This review summarizes recent progress in AI-driven polymer science, including structural representation, generative molecular design, property prediction, retrosynthetic planning, and closed-loop experimental processes. We discuss how polymer-specific encodings like BigSMILES and graph-based models enable machine-readable macromolecular structures, supporting transformer models, graph neural networks, and uncertainty-aware predictors that accurately estimate thermal, mechanical, dielectric, and functional properties. Beyond prediction, we examine generative architectures such as variational autoencoders, adversarial models, diffusion models, and reinforcement learning that enable inverse polymer design by exploring vast chemical spaces under performance constraints. Combining these computational approaches with techniques like Bayesian optimization, active learning, and autonomous labs signifies a shift toward closed-loop polymer engineering, where AI continuously suggests, synthesizes, analyzes, and refines materials with minimal human intervention. Key challenges, including data scarcity, polymer-specific representation learning, model interpretability, infrastructure integration, and sustainability goals, are thoroughly considered. Emerging solutions such as FAIR data ecosystems, physics-informed neural networks, explainable AI, and modular self-driving laboratories are evaluated for scalable implementation. Overall, these advancements point toward autonomous polymer discovery, with AI serving not only as a predictive tool but as the foundation for intelligent, adaptive, and sustainable polymer engineering.

Keywords: *Artificial Intelligence-Driven Materials Design; Polymer Informatics; Generative Polymer Design; Autonomous Laboratories; Closed-Loop Materials Discovery; Bayesian Optimization; Active Learning; Polymer Structural Representation; Self-Driving Laboratories; Sustainable Polymer Engineering.*

Introduction

Historically, the creation of polymeric materials has been time-consuming and has largely relied on empirical methods and iterative testing. This method often results in lengthy durations and significant resource utilization. The emergence of artificial intelligence (AI) and machine learning (ML) has made data-driven design, predictive modelling, and autonomous testing possible, with the potential to significantly alter polymer research.

The foundation for incorporating AI into polymer research has been established by recent developments in polymer informatics. Advancements such as the Generative BigSMILES

*Corresponding author: vagbogo@gmail.com

notation have improved the representation of complex polymer structures, enabling more precise machine-learning models for material design and property prediction. Furthermore, by capturing the sequential nature of polymer chains, transformer-based models such as TransPolymer have demonstrated greater effectiveness in forecasting polymer properties [1][2][3]. These advances reflect a broader trend across materials science, where deep learning models are increasingly employed to extract patterns from high-dimensional data and automate the prediction of material behaviour [4][5].

Significant advancements in polymer design are also being made via generative models. To tackle the problem of conformational variety in polymers, for example, polyGen, a latent diffusion model, has been created to produce realistic 3D polymer structures with little input. Similarly, PolyConf uses hierarchical generative models to generate superior polymer conformations, thereby improving polymer simulation and modelling [6]. Self-driving labs have emerged through the integration of AI into experimental workflows, automating the synthesis and characterization of polymers. This approach is best exemplified by platforms such as Polybot at Argonne National Laboratory, which uses AI-driven automation to accelerate the search for electronic polymers with ideal properties. By effectively navigating large experimental areas, these autonomous devices can significantly reduce the time and cost associated with conventional techniques [7][8].

Furthermore, reviews in this journal have highlighted how emerging materials technologies, such as nanocomposites and smart polymer interfaces, could benefit from predictive design tools and computational frameworks [9][10][11]. While most current applications remain experimental, the integration of AI promises to shift these approaches toward more scalable and systematic design strategies.

Despite these developments, there are still challenges in applying AI to polymer research. Continuous research and development are required to address challenges such as the complexity of polymer structures, data scarcity, and model interpretability. For AI to reach its full potential in this area, several issues must be resolved.

The goal of this paper is to provide a comprehensive overview of the current state of AI applications in polymer systems. We will examine the origins and development of polymer informatics, investigate the potential of generative models for polymer design, analyze how AI might be integrated into experimental workflows, and discuss the challenges and potential paths forward in this rapidly evolving field.

Polymer informatics: Foundations and progress

The use of data science and machine learning methods in polymer research, or polymer informatics, has emerged as a key component in the fusion of artificial intelligence and materials science. Because of their hierarchical architectures, polydispersity, stochastic monomer sequences, and macromolecular structures, polymers are inherently more complex than tiny molecules or inorganic crystals. Building AI-driven models for polymer prediction and design has posed several challenges, one of which is capturing this complexity in a computationally tractable manner.

A conceptual framework connecting polymer structure representation, curated data sources, machine learning models, and property prediction is shown in Figure 1 to demonstrate the fundamental elements and data flow of contemporary polymer informatics. The picture illustrates how the informatics pipeline uses a variety of encoding techniques as its first input, including SMILES, BigSMILES, and graph-based representations. Machine learning models trained to predict important polymer properties such as mechanical strength, dielectric constant, and glass transition temperature are informed by these encodings, which are fed into structured databases that may contain experimental, synthetic, or open-source datasets. The data-centric architecture that supports AI-enabled polymer discovery is best illustrated by this closed-loop pipeline.

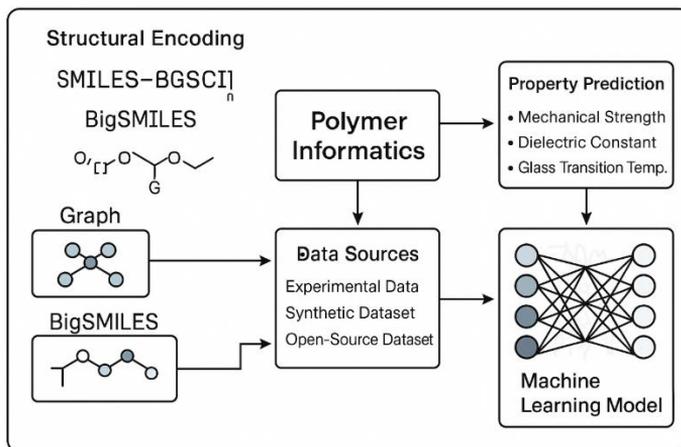


Fig. 2. Polymer Informatics Framework: From Structural Representation to Machine Learning-Based Property Prediction

Structural encoding and data representation

Polymer informatics relies on accurate and machine-readable representations of polymer structures. When applied to stochastic polymers, traditional chemical representations such as SMILES (Simplified Molecular-Input Line-Entry System) often fail, leading to the development of more specialized notations. To expand SMILES' capacity to describe polymer repeat units and their statistical connectedness, BigSMILES was created [12]. More recently, greater flexibility and compatibility with generative models have been enabled by extensions such as Generative BigSMILES [1]. In parallel, machine learning algorithms have been trained using alternative encoding techniques such as hierarchical descriptors (e.g., composition, topology, and connectedness), graph-based representations, and molecular fingerprints. These illustrations are crucial for understanding how polymer structure relates to properties such as mechanical strength, dielectric constant, and glass transition temperature (T_g) [13][14].

Data infrastructure and polymer databases

Another crucial element of polymer informatics is the availability of well-selected, excellent datasets. To aid model development, numerous efforts have been made to assemble polymer-specific databases. Among the examples are:

Experimental data on the structures and properties of polymers are available in the Japanese polymer database PolyInfo. To train and evaluate deep learning models, PI1M (Polymer Informatics 1 million) is a synthetic dataset of polymer structures [15]. A large open-source dataset called Open Macromolecular Genome was created to help generative models identify polymers [16]. These resources, along with developments in high-throughput simulations, online scraping, and literature extraction using natural language processing (NLP), are beginning to close the data gap that has historically constrained polymer informatics.

Using machine learning to predict properties

Using structural inputs to predict polymer properties has been a common application of supervised machine learning algorithms. In smaller datasets, models such as gradient-boosted decision trees (GBDTs), random forests (RFs), and support vector machines (SVMs) have demonstrated efficacy. However, by working directly on raw polymer structures or sequences, recent advances in deep learning architectures, such as transformer models and graph neural networks (GNNs), have enabled improved performance [17].

Notably, by capturing long-range interactions and chain dependencies, TransPolymer, a polymer-specific transformer architecture, has shown excellent performance in forecasting mechanical and thermal properties [2]. To direct experimental design and material optimization, these models are increasingly being combined with active learning and uncertainty quantification.

Polymer design generative models

Designing and discovering new polymeric materials has become easier because of the use of generative models in polymer research. Exploring the vast chemical space of polymers has been enabled by these models, including diffusion models, Generative Adversarial Networks (GANs), and Variational Autoencoders (VAEs).

Generative adversarial networks and variational autoencoders

By discovering the underlying distribution of pre-existing polymer datasets, VAEs, and GANs have been used to create novel polymer architectures. The inverse design process, in which desired properties guide the creation of equivalent polymer structures, is facilitated by these models. For example, GANs have been used to design polymers with certain mechanical properties, while VAEs have been used to build polymers with desired glass transition temperatures. The merits and limits of six deep generative models, VAE, AAE, ORGAN, CharRNN, REINVENT, and GraphINVENT, in de novo polymer design were assessed in benchmarking research [18].

Six popular deep generative models for polymer design are compared in Table 1, which also highlights the fundamental ideas, structural elements, and working processes of each model. These models capture different facets of polymer structure and generation dynamics, ranging from sequence-based techniques such as Variational Autoencoders (VAEs) and Character RNNs to graph-based models such as GraphINVENT. AAE adds distributional control by adversarial training, ORGAN incorporates domain-specific goals through reinforcement, and VAEs provide smooth latent spaces for interpolation. Each model architecture offers distinct benefits. While REINVENT and CharRNN offer scalable frameworks for sequence-based polymer design using SMILES or p-SMILES representations, graph-based models like GraphINVENT excel at capturing topological subtleties essential to polymer backbones. To help researchers choose the best models for specific generative polymer challenges, associated citations provide additional background information and technical details.

The fundamental topologies of six deep generative models often used in polymer design workflows are shown schematically in Figure 2. These model topologies illustrate the many approaches used to represent and produce polymer structures across various data modalities. Probabilistic encoding and reconstruction are the foundation of latent space models such as VAEs and AAEs, with the latter also using adversarial training to enforce restrictions on the latent distribution. ORGAN is particularly useful for property-optimized polymer discovery since it adds reinforcement signals to guide production toward user-defined goals. Sequence-based models that use recurrent neural networks for token-by-token generation, such as CharRNN and REINVENT, work with linearized representations like p-SMILES. On the other hand, GraphINVENT uses a graph-based methodology that enables the progressive construction of polymer structures at the atomic level. These topologies collectively demonstrate the variety of generative modelling approaches and highlight how crucial architectural choices are in shaping design integrity, diversity, and chemical validity in AI-driven polymer informatics.

Table 1. Key features and mechanisms of generative deep learning models for polymer design

Model Name	Core Concept/Goal	Key Components	How it Works	Reference
Variational Autoencoder (VAE)	Data generation and latent space learning	Encoder, Decoder	Maps input to latent space; reconstructs input; uses KL divergence and reconstruction loss	[19]
Adversarial Autoencoder (AAE)	Combines autoencoding with adversarial regularization	Encoder, Decoder, Discriminator	A latent vector is trained to follow a prior using a discriminator; regularized reconstruction	[20]
Objective-Reinforced GAN (ORGAN)	GANs guided by specific objective functions	Generator, Discriminator	Generator trained via reinforcement: Reward = $\lambda(\text{Discriminator}) + 1 - \lambda(\text{Objective})$	[21]
Character RNN (CharRNN)	Character-level sequence modeling	LSTM or GRU	Learns sequential dependencies to generate tokens like p-SMILES	[22]
REINVENT	Reinforcement learning for de novo molecule generation	Embedding, RNN layers (GRU/LSTM), Linear+Softmax	Uses a policy-based reinforcement loop with SMILES tokenization	[23]
GraphINVENT	Graph-based molecule generation	Graph Neural Network (GNN), MLP	Stepwise graph construction via learned actions: add node, edge, or stop	[24]

Although generative models have greatly improved the computational design of polymers, their full potential can only be realized when combined with experimental procedures that enable synthesis, iterative improvement, and real-world validation.

AI Integration with experimental workflows

Polymer research is becoming more autonomous, adaptive, and efficient because of the combination of automated experimentation and artificial intelligence. The pace of discovery has been constrained by the resource-intensive and sequential nature of polymer synthesis and characterization in the past. However, recent developments in closed-loop systems, self-driving labs, and active learning frameworks are now enabling the direct integration of AI-generated hypotheses into experimental execution, significantly speeding up materials development cycles.

Autonomous labs

Autonomous labs are a paradigm change in experimental polymer research. These platforms use AI decision-making, automated characterization, robotics, and high-throughput synthesis to perform experiments with little assistance from humans. Argonne National Laboratory's Polybot platform, which autonomously explores the design space of semiconducting polymers for electrical applications, is a noteworthy example [25]. Compared with manual methods, Polybot has shown notable improvements in throughput and discovery efficiency through AI-driven iterative trial planning and evaluation. By combining robots, artificial intelligence (AI), and automated systems, autonomous labs are revolutionizing experimental polymer research by enabling experiments with little human assistance. These platforms speed up the discovery and development of polymer materials by combining AI-driven decision-making, high-throughput synthesis, automated characterization, and real-time data processing.

The Polybot platform, created at Argonne National Laboratory to independently explore the design space of semiconducting polymers for electrical applications, is a well-known example. When compared to conventional manual approaches, Polybot has shown notable improvements in throughput and discovery efficiency through the iterative planning and evaluation of trials utilizing AI algorithms [26]. In addition to Polybot, significant progress has been made in self-driving laboratories (SDLs) for materials science and chemistry.

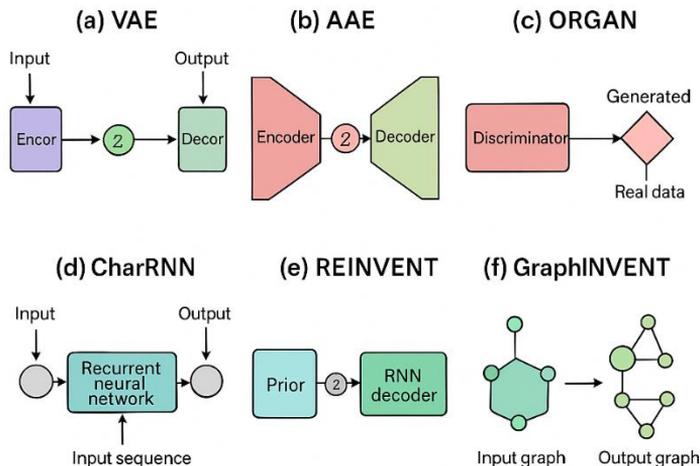


Fig. 3. Polymer design procedures commonly use deep generative model topologies, which highlight significant variations in representation and generation techniques: (a) variational autoencoder (VAE), (b) adversarial autoencoder (AAE), (c) objective-reinforced generative adversarial network (ORGAN), (d) character recurrent neural network (CharRNN), (e) REINVENT, and (f) GraphINVENT

These SDLs coordinate intricate experimental workflows by combining automated hardware and advanced software. Modular robotic platforms that perform multi-step syntheses, enable real-time monitoring using analytical methods such as FT-IR and HPLC, and support adaptive decision-making using machine learning algorithms are recent advancements. Three main parts usually make up the software architecture of SDLs: (1) control systems for automated hardware, (2) tools for data extraction, management, and analysis, and (3) AI-powered experimental planners. This integration greatly reduces the time and resources needed for polymer discovery by enabling data-driven optimization of reaction conditions, quick iteration cycles, and autonomous experiment execution [26].

Bayesian optimization and active learning

By prioritizing the most informative data points, active learning algorithms are increasingly used to guide experiments. These methods work especially well in polymer systems where the vastness of the design space makes comprehensive sampling impossible. For instance, stretchy polymers and block copolymer morphologies have been effectively discovered using Bayesian optimization, which maximizes performance metrics with fewer experimental trials [27]. A potent closed-loop system that repeatedly develops, tests, and improves hypotheses in real time is produced by combining such tactics with generative models. Metrics such as model uncertainty, variance reduction, or predicted improvement are frequently used in active learning procedures to identify data points with the greatest potential for knowledge gain. To improve model performance with the least experimental expense, these points are prioritized for labelling or experimentation. In polymer informatics, where experimental synthesis and characterization can be costly and time-consuming, this approach is highly beneficial.

A particular type of active learning called Bayesian optimization (BO) has become well known for its ability to optimize black-box functions with few evaluations. Rohr et al. [28], for instance, showed how to use BO to tune the performance of polymer membranes by repeatedly proposing new candidates with the highest expected property improvement, significantly lowering the number of experimental cycles [28]. To capture intricate, non-linear correlations between polymer structure and performance, BO has been further merged with deep kernel

learning or Gaussian processes. It has been demonstrated that BO can efficiently traverse formulation spaces for thermoset composites and ionic conductors in high-throughput experimentation setups [29][30].

Furthermore, uncertainty-aware generative models have been combined with active learning to enable iterative sampling from model-recommended choices with measurable confidence levels. This method serves as the basis for autonomous closed-loop discovery systems, in which artificial intelligence (AI) agents can both suggest novel polymers and evaluate the accuracy of their predictions before initiating synthesis. Like frameworks are demonstrated by tools like Gryffin and ChemOS, which combine intelligent sampling with design-of-experiments to optimize chemical and polymer operations [31]. These methods demonstrate the importance of combining AI-driven design with intelligent sampling algorithms in polymer science, as this approach greatly increases efficiency, lowers experimental costs, and improves the likelihood of finding materials with unique and ideal specifications.

Retrosynthetic planning and process automation

AI facilitates retrosynthetic analysis and process design in addition to property prediction and candidate generation. While deep reinforcement learning models have been used to predict possible reaction sequences for complicated polymer targets, natural language processing methods have been utilized to extract polymer synthesis pathways from the literature [32]. These techniques minimize the gap between theory and experiment by providing an integrated pathway from in silico design to benchtop synthesis when combined with robotic automation. Although tiny molecules have traditionally been the focus of retrosynthetic planning, current initiatives have broadened these techniques to account for the complexity of polymer systems. Large, multifunctional monomers and step-growth polymerization processes have led to the adoption of retrosynthesis models such as ASKCOS [33], which uses neural-symbolic AI and template-based AI to map backward from desired polymers to readily accessible monomers. These models propose viable synthetic routes based on cost, step count, or environmental variables, using chemical reaction rules and databases such as Reaxys or PubChem [32]. In the meantime, NLP-based tools such as PolySynthesis and ChemDataExtractor automatically search the literature for catalysts, solvents, polymerization settings, and yields. These methods are increasingly used to select high-quality synthesis datasets for training machine learning models. For instance, Guo et al., [34] developed a sequence-to-sequence model for polymer process planning using an NLP-extracted dataset of more than 4,000 polymerization processes [34].

Polymer retrosynthesis has also seen the adoption of deep learning frameworks such as Molecular Transformer, in which attention-based architectures can learn context-sensitive reaction patterns without the need for manually created rules. In predicting regioselective and stereoselective transformations, which are crucial characteristics in polymer synthesis, this model has been demonstrated to perform better than conventional rule-based systems [35].

These planning tools enable the autonomous execution of multistep polymer synthesis routes when combined with robotic platforms (such as flow reactors or liquid-handling robots). Such integration is demonstrated by projects such as IBM RoboRXN and the Cronin Lab's Chemputer, which closes the loop from digital design to physical manifestation by translating AI-predicted routes directly into machine-readable scripts for synthesis execution, verification, and refinement [36][37]. When combined, these advancements in automation and AI-assisted retrosynthesis provide a strong basis for intelligent polymer design pipelines that minimize human labour while increasing synthesis efficiency and repeatability.

A closed-loop design framework, in which every step of the polymer discovery pipeline feeds into the next to enable continual development, provides the best understanding of how AI can be integrated with automated experimentation. This iterative cycle starts with a generative model that suggests candidate polymers based on desired property profiles, as shown in Figure 3. Platforms for automated synthesis and characterization are used to carry out the experimental

design, which is informed by these candidates. The generative model is then improved based on feedback from evaluating and validating the resulting experimental data. The features of next-generation polymer research, accelerated hypothesis generation, effective resource utilization, and adaptive decision-making, are made possible by this AI-coordinated loop.

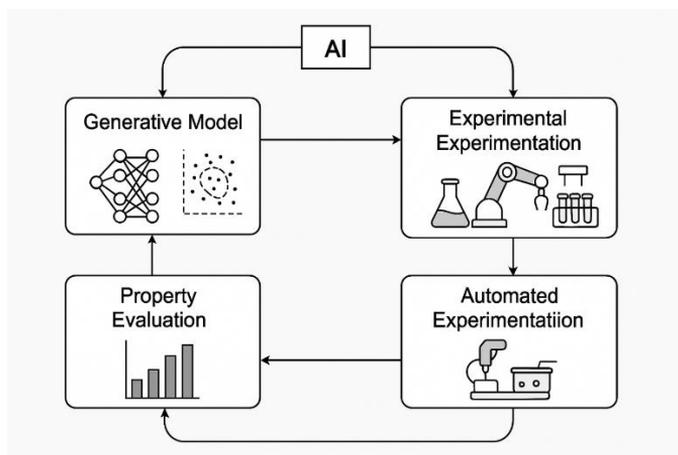


Fig. 4. End-to-end AI-integrated workflow for intelligent polymer design and experimental execution

Challenges and future directions

Despite the quick development of AI's use in polymer science, several important challenges remain before the technology's transformative potential can be completely realized. These issues span the technological, methodological, and practical realms and will require interdisciplinary collaboration, strong infrastructure, and ongoing innovation.

Many AI-driven polymer systems still have limitations in breadth, scalability, and reliability when used in real-world research and industry, despite remarkable advances in automation, generative design, and property prediction. These restrictions stem from fundamental problems, including incomplete data ecosystems, underdeveloped structural representations, and difficult-to-interpret predictive models. However, as models begin to offer viable candidates for synthesis, ethical considerations related to sustainability and material safety are becoming increasingly significant, and the technical integration of AI into experimental pipelines remains challenging.

This section describes the main systemic and technical obstacles to the wider use of AI in polymer science, along with potential future paths to turn today's promising prototypes into reliable, scalable solutions. For next-generation polymer discovery to have a coherent, intelligent, and sustainable infrastructure, these obstacles must be removed.

Challenges

Quality and scarcity of data

The absence of substantial, consistent, and high-quality datasets is one of the most enduring challenges in polymer informatics. Polymers have a far greater degree of structural complexity than tiny molecules or crystalline materials, including polydispersity, stochastic chain sequences, and hierarchical organization. Most publicly accessible datasets have biased sampling, insufficient experimental annotations, or inconsistent representations. One partial solution is the generation of synthetic data through simulations or transfer learning from better-annotated domains [38][39]. However, the field urgently needs community efforts toward open, FAIR (Findable, Accessible, Interoperable, Reusable) polymer databases.

Complex polymer architectures: representation of learning

It is still unclear how to encode polymer structures in a way that is both chemically accurate and machine-readable. Representing branched, block, and cross-linked architectures remains challenging, despite recent advancements such as BigSMILES, SELFIES, and graph-based representations that improve the quality of model input. Future research must focus on hybrid and hierarchical representations that can capture macromolecular architecture and monomer-level chemistry simultaneously. This gap might be partially filled by new techniques such as physically informed embeddings and 3D graph neural networks.

Interpretability and credibility of the model

Deep learning models are frequently black boxes, especially when it comes to generative design or property prediction. Interpretability and model reliability are critical in high-stakes fields such as biomedical polymers, aircraft coatings, and recyclable plastics. While methods like counterfactual analysis, attention-weight visualizations, and SHAP (SHapley Additive exPlanations) can offer some partial insights, additional effort is required to create interpretable-by-design models. In polymer informatics, explainable AI (XAI) will be essential to enabling uptake in both industrial and experimental contexts [40][41][42].

Autonomous system integration and scalability

Although the AI-polymer pipeline's separate elements, generative models, retrosynthesis tools, and self-driving labs, have demonstrated potential, large-scale integration is still difficult. Interoperable platforms, reliable automation hardware, and advanced orchestration software are necessary to seamlessly integrate in silico design with real-world synthesis, validation, and feedback. Particularly in multi-step or multi-material workflows, scalable solutions must also address concerns such as reagent compatibility, synthesis time, and real-time data assimilation [43].

Sustainable and ethical AI in material design

The importance of ethical considerations will increase as AI plays a bigger role in materials innovation. This covers the environmental effects of AI-recommended polymer materials, including their toxicity, biodegradability, and circularity. Environmental and life-cycle factors ought to be incorporated into future models as optimization goals. Furthermore, to align AI research with green chemistry objectives, sustainable computing techniques such as low-resource inference and energy-efficient model training should be prioritized [44].

Prospects

Combining automation, artificial intelligence, and polymer chemistry may one day open up hitherto unimaginable avenues for the discovery of novel materials. Soon, we expect to see more generalizable polymer representations, experimental AI platforms, and models that not only generate novel polymer structures but also validate and evaluate them. Equally important will be the democratization of these technologies through open-source platforms, community-curated databases, and equitable access to high-throughput equipment. If these challenges are collectively resolved, next-generation AI will transform polymer research from an art based on intuition into a data-centric, predictive, and autonomous field.

Looking ahead, the emergence of self-driving labs designed especially for polymer research will be one of the most revolutionary advances. With the help of closed-loop AI algorithms that optimize synthesis and characterization in real time, these labs will allow for autonomous experimentation. Leading systems have already demonstrated that AI-guided feedback loops can explore large compositional design spaces, identify new catalysts, and optimize polymer processing conditions at speeds not possible with conventional methods [31][36].

On the computational front, domain-aware structures that account for polymer connectivity, conformation, and dynamics, as well as physics-informed neural networks, will be incorporated into polymer-specific generative models as they further develop. To replicate real-world

behaviors like degradation, crystallization, self-assembly, and recyclability parameters that are becoming increasingly important for sustainability and circular materials design, these models will probably go beyond static property prediction [6][45].

The global movement toward open-access data ecosystems and the standardization of polymer data formats will be equally significant. Reducing data silos and improving the discoverability and reusability of polymer datasets are the goals of FAIR-aligned projects like the Materials Genome Initiative and the Open Macromolecular Genome [6][15]. The benchmarking and validation of AI models will be enabled by these community-driven resources, which will also encourage reproducibility across organizations and sectors [45-49].

Additionally, modular design tools and cloud-based AI platforms are expected to democratize access to cutting-edge capabilities. User-friendly AI interfaces for polymer design could enable non-experts, such as formulation chemists and product designers, to collaborate to develop new materials using intelligent, intuitive algorithms, much like AlphaFold has revolutionized protein science.

Lastly, the development of AI in polymer science will be heavily reliant on sustainable and ethical innovation. Next-generation tools must incorporate environmental and lifespan considerations alongside technical performance. Examples include reducing toxicity, improving recyclability, and lowering energy costs during model training. Only through this comprehensive integration can AI both accelerate discovery and align polymer development with global objectives.

Conclusion

The field of polymer research is rapidly changing due to artificial intelligence, which offers previously unheard-of possibilities to improve design accuracy, accelerate discovery, and automate experimental procedures. This research examines the integration of next-generation AI techniques throughout the polymer development process, spanning autonomous synthesis and retrosynthetic planning, data-driven property prediction, and generative modeling. Chemists can now traverse the vast chemical space of polymers more effectively and purposefully thanks to advances in polymer informatics, deep generative models, and active learning frameworks.

However, several obstacles still exist despite these successes. The scalability and practical implementation of these technologies are still hindered by a lack of data, inconsistent chemical representations, limited model interpretability, and the technical challenges of integrating AI with laboratory equipment. A coordinated effort by chemists, computer scientists, automation engineers, and legislators will be needed to overcome these constraints.

In the future, the combination of AI, automation, and sustainable design principles will change how we think about and conduct polymer research and accelerate the development of new materials. The discipline will shift firmly from intuition-led workflows to a predictive, data-centric paradigm as polymer-specific databases expand, generative models gain chemical awareness, and self-driving labs emerge.

For the advantages of AI-driven polymer discovery to be widely disseminated and used ethically, these technologies must be made more accessible through open-source software, publicly available data platforms, and collaborative infrastructure. Next-generation AI has the potential to usher in a new era of intelligent polymer design that is faster, greener, and more innovative than ever before, if it is supported by sustained investment and strategic vision.

Declaration of competitive interest

The authors state that any known competing financial interests or personal relationships could not have influenced the work described in this study.

Credit authorship contribution statement

Dr. Agbogo: Writing – writing – original draft, conceptualization. Prof. Sadiku: review, Supervision. Dr. Mavhungu: Project administration and investigation. Dr. Teffo: Validation, Editing, and Dr. Injor: Visualization.

Acknowledgments

This research was supported by Tshwane University of Technology (TUT), Pretoria, South Africa. We certify that the manuscript represents our original work, that all data are authentic and reproducible, and that no AI system was used to fabricate data, results, or interpretations. We only used Grammarly to enhance the readability and language of the manuscript. Thank you for the opportunity to publish with your esteemed journal.

References

- [1] Schneider, L., Walsh, D., Olsen, B., & de Pablo, J. (2024). Generative BigSMILES: an extension for polymer informatics, computer simulations & ML/AI. *Digital Discovery*, 3(1), 51–61. <https://doi.org/10.1039/D3DD00147D>
- [2] Xu, C., Wang, Y., & Barati Farimani, A. (2023). TransPolymer: a Transformer-based language model for polymer property predictions. *Npj Computational Materials*, 9(1), 64. <https://doi.org/10.1038/s41524-023-01016-5>
- [3] Gruzicic, M., Ramaswami, S., S. Snipes, J., & Yavari, R. (2014). Multi-Scale Computation-Based Design of Nano-Segregated Polyurea for Maximum Shockwave-Mitigation Performance. *AIMS Materials Science*, 1(1), 15–27. <https://doi.org/10.3934/mat.2014.1.15>
- [4] Martin, T. B., & Audus, D. J. (2023). Emerging Trends in Machine Learning: A Polymer Perspective. *ACS Polymers Au*, 3(3), 239–258. <https://doi.org/10.1021/acspolymersau.2c00053>
- [5] Struble, D. C., Lamb, B. G., & Ma, B. (2024). A prospective on machine learning challenges, progress, and potential in polymer science. *MRS Communications*, 14(5), 752–770. <https://doi.org/10.1557/s43579-024-00587-8>
- [6] Ayush Jain, & Rampi Ramprasad. (2025). polyGen: A Learning Framework for Atomic-level Polymer Structure Generation. *Computer Science > Computational Engineering, Finance, and Science*. <https://doi.org/https://doi.org/10.48550/arXiv.2504.17656>
- [7] Amber Rose. (2025, February). *Self-driving lab transforms materials discovery Harnessing the power of artificial intelligence and automation to uncover new possibilities*. Argonne National Laboratory. https://doi.org/https://www.anl.gov/article/selfdriving-lab-transforms-materials-discovery?utm_source=.com
- [8] Vriza, A., Chan, H., & Xu, J. (2023). Self-Driving Laboratory for Polymer Electronics. *Chemistry of Materials*, 35(8), 3046–3056. <https://doi.org/10.1021/acs.chemmater.2c03593>
- [9] Vilvanatha Prabu A., Vijayaraghavan G.V, Suriakarthick R. & Priscilla J. (2025). Investigation on Structural, Morphological, and Electrical Properties of Novel PVDF/ZnSb Polymer Nanocomposites for Flexible Energy Storage Applications. *Polymer-Plastics Technology and Materials*, 1–12. <https://doi.org/10.1080/25740881.2025.2475079>
- [10] Sabti, H. M. (2025). Enhanced Properties of Aramid/Salinized MWCNT Composites Through Surface Treatment: A Combined Chemical and Microscopic Approach. *Polymer-Plastics Technology and Materials*, 1–20. <https://doi.org/10.1080/25740881.2025.2477106>
- [11] Agbogo, V. U., Sadiku, E. R., Mavhungu, L., Kupolati, W. K., & Injor, O. M. (2025). Nanotechnology coatings in the defense and aerospace industry. *Next Nanotechnology*, 7, 100197. <https://doi.org/10.1016/j.nxnano.2025.100197>
- [12] Lin, T.-S., Coley, C. W., Mochigase, H., Beech, H. K., Wang, W., Wang, Z., Woods, E., Craig, S. L., Johnson, J. A., Kalow, J. A., Jensen, K. F., & Olsen, B. D. (2019). BigSMILES: A Structurally-

- Based Line Notation for Describing Macromolecules. *ACS Central Science*, 5(9), 1523–1531. <https://doi.org/10.1021/acscentsci.9b00476>
- [13] Bal, H., & Pal, A. (2020). Parallel and Distributed Machine Learning Algorithms for Scalable Big Data Analytics. *Future Generation Computer Systems*, 108, 1159–1161. <https://doi.org/10.1016/j.future.2019.07.009>
- [14] Upadhyaya, S. R. (2013). Parallel approaches to machine learning—A comprehensive survey. *Journal of Parallel and Distributed Computing*, 73(3), 284–292. <https://doi.org/10.1016/j.jpdc.2012.11.001>
- [15] Kim, S., Schroeder, C. M., & Jackson, N. E. (2023). Open Macromolecular Genome: Generative Design of Synthetically Accessible Polymers. *ACS Polymers Au*, 3(4), 318–330. <https://doi.org/10.1021/acspolymersau.3c00003>
- [16] Chen, L., Pilania, G., Batra, R., Huan, T. D., Kim, C., Kuenneth, C., & Ramprasad, R. (2021). Polymer informatics: Current status and critical next steps. *Materials Science and Engineering: R: Reports*, 144, 100595. <https://doi.org/10.1016/j.mser.2020.100595>
- [17] Choudhary, K., DeCost, B., Chen, C., Jain, A., Tavazza, F., Cohn, R., Park, C. W., Choudhary, A., Agrawal, A., Billinge, S. J. L., Holm, E., Ong, S. P., & Wolverton, C. (2022). Recent advances and applications of deep learning methods in materials science. *Npj Computational Materials*, 8(1), 59. <https://doi.org/10.1038/s41524-022-00734-6>
- [18] Yue, T., Tao, L., Varshney, V., & Li, Y. (2025). Benchmarking study of deep generative models for inverse polymer design. *Digital Discovery*, 4(4), 910–926. <https://doi.org/10.1039/D4DD00395K>
- [19] Gómez-Bombarelli, R., Wei, J. N., Duvenaud, D., Hernández-Lobato, J. M., Sánchez-Lengeling, B., Sheberla, D., Aguilera-Iparraguirre, J., Hirzel, T. D., Adams, R. P., & Aspuru-Guzik, A. (2018). Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Central Science*, 4(2), 268–276. <https://doi.org/10.1021/acscentsci.7b00572>
- [20] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, & Ian Goodfellow Brendan Frey. (2015). Adversarial Autoencoders. *Computer Science > Machine Learning*. <https://doi.org/https://doi.org/10.48550/arXiv.1511.05644>
- [21] Guimaraes, G. L., Sanchez-Lengeling, B., Outeiral, C., Farias, P. L. C., & Aspuru-Guzik, A. (2017). *Objective-reinforced generative Adversarial Networks (ORGAN) for Sequence Generation Models*. <https://doi.org/10.48550/arXiv.1705.10843>
- [22] Arús-Pous, J., Johansson, S. V., Prykhodko, O., Bjerrum, E. J., Tyrchan, C., Reymond, J.-L., Chen, H., & Engkvist, O. (2019). Randomized SMILES strings improve the quality of molecular generative models. *Journal of Cheminformatics*, 11(1), 71. <https://doi.org/10.1186/s13321-019-0393-0>
- [23] Olivecrona, M., Blaschke, T., Engkvist, O., & Chen, H. (2017). Molecular de-novo design through deep reinforcement learning. *Journal of Cheminformatics*, 9(1), 48. <https://doi.org/10.1186/s13321-017-0235-x>
- [24] Mercado, R., Rastemo, T., Lindelöf, E., Klambauer, G., Engkvist, O., Chen, H., & Jannik Bjerrum, E. (2021). Graph networks for molecular design. *Machine Learning: Science and Technology*, 2(2), 025023. <https://doi.org/10.1088/2632-2153/abc91>
- [25] Joseph E. Harmon. (2023, April 25). *Argonne's self-driving lab accelerates the discovery process for materials with multiple applications: Autonomous discovery lab leading the way in transforming scientific research on sustainable and bio-inspired microelectronics*. Argonne National Laboratory. <https://doi.org/https://www.anl.gov/article/argonnes-selfdriving-lab-accelerates-the-discovery-process-for-materials-with-multiple-applications>
- [26] Tom, G., Schmid, S. P., Baird, S. G., Cao, Y., Darvish, K., Hao, H., Lo, S., Pablo-García, S., Rajaonson, E. M., Skreta, M., Yoshikawa, N., Corapi, S., Akkoc, G. D., Strieth-Kalthoff, F., Seifrid, M., & Aspuru-Guzik, A. (2024). Self-Driving Laboratories for Chemistry and Materials Science. *Chemical Reviews*, 124(16), 9633–9732. <https://doi.org/10.1021/acs.chemrev.4c00055>

- [27] Zhang, Y., Apley, D. W., & Chen, W. (2020). Bayesian Optimization for Materials Design with Mixed Quantitative and Qualitative Variables. *Scientific Reports*, 10(1), 4924. <https://doi.org/10.1038/s41598-020-60652-9>
- [28] Rohr, B., Stein, H. S., Guevarra, D., Wang, Y., Haber, J. A., Aykol, M., Suram, S. K., & Gregoire, J. M. (2020). Benchmarking the acceleration of materials discovery by sequential learning. *Chemical Science*, 11(10), 2696–2706. <https://doi.org/10.1039/C9SC05999G>
- [29] Batra, R., Song, L., & Ramprasad, R. (2020). Emerging materials intelligence ecosystems propelled by machine learning. *Nature Reviews Materials*, 6(8), 655–678. <https://doi.org/10.1038/s41578-020-00255-y>
- [30] Günay, M. E., & Tapan, N. A. (2023). Evaluation of polymer electrolyte membrane electrolysis by explainable machine learning, optimum classification model, and active learning. *Journal of Applied Electrochemistry*, 53(3), 415–433. <https://doi.org/10.1007/s10800-022-01786-8>
- [31] Häse, F., Roch, L. M., & Aspuru-Guzik, A. (2019). Next-Generation Experimentation with Self-Driving Laboratories. *Trends in Chemistry*, 1(3), 282–291. <https://doi.org/10.1016/j.trechm.2019.02.007>
- [32] Coley, C. W., Rogers, L., Green, W. H., & Jensen, K. F. (2017). Computer-Assisted Retrosynthesis Based on Molecular Similarity. *ACS Central Science*, 3(12), 1237–1245. <https://doi.org/10.1021/acscentsci.7b00355>
- [33] Zhengkai Tu, Sourabh J. Choure, Mun Hong Fong, Jihye Roh, Itai Levin, Kevin Yu, Joonyoung F. Joung, Nathan Morgan, Shih-Cheng Li, Xiaoqi Sun, Huiqian Lin, Mark Murnin, Jordan P. Liles, Thomas J. Struble, Michael E. Fortunato, Mengjie Liu, William H. Green, Klavs F. Jensen, & Connor W. Coley. (2025). ASKCOS: an open source software suite for synthesis planning. *Computer Science > Artificial Intelligence*. <https://doi.org/https://doi.org/10.48550/arXiv.2501.01835>
- [34] Guo, J., Ibanez-Lopez, A. S., Gao, H., Quach, V., Coley, C. W., Jensen, K. F., & Barzilay, R. (2022). Automated Chemical Reaction Extraction from Scientific Literature. *Journal of Chemical Information and Modeling*, 62(9), 2035–2045. <https://doi.org/10.1021/acs.jcim.1c00284>
- [35] Schwaller, P., Laino, T., Gaudin, T., Bolgar, P., Hunter, C. A., Bekas, C., & Lee, A. A. (2019). Molecular Transformer: A Model for Uncertainty-Calibrated Chemical Reaction Prediction. *ACS Central Science*, 5(9), 1572–1583. <https://doi.org/10.1021/acscentsci.9b00576>
- [36] Burger, B., Maffettone, P. M., Gusev, V. v., Aitchison, C. M., Bai, Y., Wang, X., Li, X., Alston, B. M., Li, B., Clowes, R., Rankin, N., Harris, B., Sprick, R. S., & Cooper, A. I. (2020). A mobile robotic chemist. *Nature*, 583(7815), 237–241. <https://doi.org/10.1038/s41586-020-2442-2>
- [37] Steiner, S., Wolf, J., Glatzel, S., Andreou, A., Granda, J. M., Keenan, G., Hinkley, T., Aragon-Camarasa, G., Kitson, P. J., Angelone, D., & Cronin, L. (2019). Organic synthesis in a modular robotic system driven by a chemical programming language. *Science*, 363(6423). <https://doi.org/10.1126/science.aav2211>
- [38] Kim, C., Batra, R., Chen, L., Tran, H., & Ramprasad, R. (2021). Polymer design using genetic algorithm and machine learning. *Computational Materials Science*, 186, 110067. <https://doi.org/10.1016/j.commatsci.2020.110067>
- [39] Shetty, P., & Ramprasad, R. (2021). Automated knowledge extraction from polymer literature using natural language processing. *iScience*, 24(1), 101922. <https://doi.org/10.1016/j.isci.2020.101922>
- [40] Pyzer-Knapp, E. O., Pitera, J. W., Staar, P. W. J., Takeda, S., Laino, T., Sanders, D. P., Sexton, J., Smith, J. R., & Curioni, A. (2022). Accelerating materials discovery using artificial intelligence, high-performance computing, and robotics. *Npj Computational Materials*, 8(1), 84. <https://doi.org/10.1038/s41524-022-00765-z>
- [41] Rathi, P., Jha, M. K., Hata, K., & Subramaniam, C. (2017). Real-time, Wearable, Biomechanical Movement Capture of Both Humans and Robots with Metal-Free Electrodes. *ACS Omega*, 2(8), 4132–4142. <https://doi.org/10.1021/acsomega.7b00491>

- [42] Tran, H., Gurnani, R., Kim, C., Pilania, G., Kwon, H.-K., Lively, R. P., & Ramprasad, R. (2024). Design of functional and sustainable polymers assisted by artificial intelligence. *Nature Reviews Materials*, 9(12), 866–886. <https://doi.org/10.1038/s41578-024-00708-8>
- [43] Roch, L. M., Häse, F., Kreisbeck, C., Tamayo-Mendoza, T., Yunker, L. P. E., Hein, J. E., & Aspuru-Guzik, A. (2020). ChemOS: An orchestration software to democratize autonomous discovery. *PLOS ONE*, 15(4), e0229862. <https://doi.org/10.1371/journal.pone.0229862>
- [44] Pilania, G. (2021). Machine learning in materials science: From explainable predictions to autonomous design. *Computational Materials Science*, 193, 110360. <https://doi.org/10.1016/j.commatsci.2021.110360>
- [45] Çaldağ, M. T., & Gökalp, E. (2025). Understanding barriers affecting the adoption and usage of open access data in the context of organizations. *Data and Information Management*, 9(1), 100049. <https://doi.org/10.1016/j.dim.2023.100049>
- [46] Green J.S., Hart P.W., Jamison A.J., & Jensen, K. F. (2019). A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science*, 365(6453). <https://doi.org/10.1126/science.aax1566>
- [47] Runeson, P., Olsson, T., & Linåker, J. (2021). Open Data Ecosystems — An empirical investigation into an emerging industry collaboration concept. *Journal of Systems and Software*, 182, 111088. <https://doi.org/10.1016/j.jss.2021.111088>
- [48] Agbogo, V. U., Sadiku, E. R., Mavhungu, L. M., & Teffo, M. L. (2025). Next-generation biodegradable polymers: toward a circular plastics economy. *AIMS Bioengineering*, 12(4), 473–502. <https://doi.org/10.3934/bioeng.2025023>
- [49] Agbogo, V. U., Sadiku, E. R., Mavhungu, L., Kupolati, W. K., Teffo, M. L., & Injor, O. M. (2026). Leveraging nanopolymer bio-coatings for biodegradable implant systems: An overview. *Results in Materials*, 29, 100868. <https://doi.org/10.1016/j.rinma.2025.100868>
-

Received: August 21, 2025

Accepted: November 05, 2025