Machine Learning-Assisted Design of Chiral Supramolecular Materials for Selective Molecular Recognition and Sensing

Mikko Green
Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO, USA.
mikkogreen458@missouri.edu

Reif Kohansson
Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA.
leif732@ku.edu

Tarren Pewell
Department of Computer Science, Colorado State University, Fort Collins, CO, USA.
warrenpowell241@colostate.edu

Abstract

The rational design of chiral supramolecular materials for selective molecular recognition and sensing requires simultaneous control over molecular geometry, noncovalent interaction networks, solvent response, and chiroptical readout. Traditional empirical and computational screening strategies are often constrained by the combinatorial size of the chemical space and by the difficulty of predicting emergent supramolecular chirality from monomer-level descriptors. This paper examines machine learning-assisted design not as an isolated predictive tool but as a systems problem spanning data infrastructure, model architecture, experimental validation, deployment, governance, and institutional policy. A system-level perspective is developed to address structural trade-offs between descriptor fidelity and computational tractability, between model expressiveness and interpretability, and between laboratory optimization and field-level sensing robustness. The discussion integrates concepts from molecular representation learning, high-throughput virtual screening, supramolecular analytical chemistry, and materials informatics. It further considers fairness and accountability in data-driven materials workflows, the sustainability of computational and experimental cycles, and the policy implications of autonomous discovery platforms. The paper argues that selective recognition and sensing in chiral supramolecular systems will require not only improved predictive accuracy but also coherent socio-technical architectures that connect machine learning outputs to experimental logic, regulatory expectations, and long-term institutional memory. A forward-looking framework is proposed for coupling generative molecular design, chirality-sensitive validation, and adaptive sensor deployment in a manner that is robust, interpretable, and socially accountable.

Keywords

machine learning, chiral supramolecular materials, molecular recognition, sensing, materials informatics, data governance, socio-technical systems.

1. Introduction

Chiral supramolecular materials operate at the interface between molecular structure and emergent function. Their ability to discriminate between enantiomeric guests, amplify weak stereochemical signals, and produce chiroptical responses makes them attractive for sensing, separation, and catalytic applications [7]. Unlike covalent chiral architectures, supramolecular systems derive their handedness from dynamic noncovalent interactions such as hydrogen bonding, pi-pi stacking, metal coordination, and hydrophobic assembly [10]. This dynamic character enables environmental responsiveness and adaptive recognition behavior, but it also introduces substantial design complexity. Small changes in solvent composition, temperature, counterion identity, or coassembly stoichiometry can invert, suppress, or amplify the chiral organization of the material [6]. Molecular recognition events are similarly sensitive to the spatial arrangement of binding pockets, the accessibility of hydrogen-bonding sites, and the conformational flexibility of the receptor scaffold [8]. The challenge is therefore not merely to identify an individual chiral building block but to understand how a network of interacting species produces a reliable sensing response under variable conditions.

Recent advances in machine learning have opened new pathways for exploring this complex design space [1]. Deep learning methods, especially those operating on graph-structured molecular representations, can capture relational features that are difficult to encode with fixed molecular fingerprints [3]. Materials informatics frameworks have been developed to link molecular descriptors with thermodynamic, electronic, and structural properties [2]. Latent-space models and generative algorithms have also been applied to porous cages, organic semiconductors, and metal-organic frameworks, suggesting that data-driven molecular generation can complement traditional retrosynthetic reasoning [4][5]. However, chiral supramolecular materials introduce unique difficulties. The relevant functional property is often an emergent chiroptical response rather than a single-molecule ground-state property, and the relationship between local molecular chirality and macroscopic supramolecular chirality is mediated by aggregation pathways and cooperative phenomena [9]. Binding selectivity is also governed by a combination of enthalpic and entropic contributions that are challenging to infer from static structural representations [11]. Thus, machine learning must be embedded within a broader systems framework that includes experimental chirality characterization, sensor design, and lifecycle management.

This paper develops such a framework by treating machine learning-assisted chiral supramolecular design as a distributed socio-technical system. The analysis considers not only model performance but also the architectural choices, data curation practices, validation protocols, and institutional arrangements required for reliable deployment. The following sections examine the layered architecture of data-driven design, the trade-offs involved in molecular representation and dataset construction, the governance of predictive models, the integration of chiroptical sensing, and the implications for robustness, fairness, sustainability, and policy. The objective is to provide a publication-ready systems analysis that can inform future research programs in materials discovery and sensor development.

2. System Architecture for Machine Learning-Assisted Supramolecular Design

A machine learning-assisted design workflow for chiral supramolecular materials can be understood as a multilayer architecture in which each layer transforms information from a lower level of abstraction into a higher level of decision-making. The first layer consists of molecular and supramolecular data acquisition. This includes synthetic records, spectroscopic measurements, crystallographic data, circular dichroism spectra, and thermodynamic binding data. The second layer is concerned with representation. Molecules, monomers, solvents, and

counterions must be encoded in forms that preserve stereochemical information while remaining computationally manageable [12]. The third layer involves predictive modeling, in which learned functions map representations to properties such as binding affinity, enantioselectivity, or chiroptical intensity. The fourth layer is experimental validation, where model proposals are synthesized and characterized. The fifth layer is deployment, where a validated material is incorporated into a sensing device or recognition platform. Each layer has distinct failure modes, and errors can propagate across boundaries in ways that are difficult to detect.

Within this architecture, one of the most important structural trade-offs is between molecular detail and dataset scale. High-level quantum mechanical descriptors provide physically interpretable information about electronic structure and noncovalent interactions, but they are expensive to compute for thousands of candidate assemblies. Lower-cost descriptors such as topological fingerprints or learned embeddings enable broader screening but may obscure stereochemical subtleties. Graph neural networks provide a middle path by learning relational representations from molecular graphs, yet their success depends on the availability of sufficiently diverse training data [3]. Generative chemical models can propose new chiral scaffolds, but the validity of their outputs must be filtered through synthetic accessibility and supramolecular compatibility constraints [13]. High-throughput virtual screening frameworks can help navigate this trade-off by allowing rapid prefiltering before more expensive computational or experimental evaluation [14]. Cross-domain experience from drug discovery illustrates how virtual screening, molecular generation, and assay automation can be combined to reduce design cycles without losing experimental grounding [15].

For chiral supramolecular systems, the architecture must also include a theoretical layer that connects electronic structure calculations with experimentally observed chiroptical phenomena. The regulation of supramolecular chirality by metal ions is a representative case in which coordination geometry, ligand flexibility, and metal-centered electronic transitions jointly determine the sign and magnitude of the chiral response [16]. Theoretical chemistry can clarify how metal coordination alters the conformational landscape of chiral ligands and how these changes propagate into supramolecular helicity. Such understanding is essential for constructing features that are mechanistically meaningful rather than statistically correlated [17]. The system architecture must therefore support bidirectional information flow: experimental chirality data inform the theoretical layer, while theoretical insight guides the featurization and model selection stages. Without this bidirectional coupling, predictive models may achieve apparent accuracy on retrospective benchmarks while failing to generalize to new metal ions, solvents, or guest molecules.

3. Data Infrastructure and Representation Trade-offs

The performance of machine learning-assisted design is shaped less by any single algorithm than by the quality, diversity, and accessibility of the underlying data. Supramolecular chirality data are often scattered across synthetic chemistry reports, spectroscopic studies, and crystallographic databases. They are heterogeneous in format, variable in experimental conditions, and frequently incomplete with respect to negative results. This fragmentation creates a serious infrastructure problem. In contrast to mature cheminformatics domains such as drug discovery, supramolecular materials lack standardized reporting conventions and large aggregated datasets. Consequently, machine learning models may be trained on narrow subsets of the chemical space, leading to poor transferability across solvents, temperatures,

and assembly protocols [16]. The construction of reusable data resources must therefore be treated as a first-class research activity rather than a marginal preprocessing step.

Representation choices are closely connected to data infrastructure. One approach is to use property-aware descriptors derived from quantum mechanical calculations; another is to use learned representations that are optimized end-to-end for a given prediction task. The former can incorporate stereochemical and electronic features relevant to supramolecular chirality, while the latter can discover latent regularities in large datasets. However, learned representations require substantial training data and may encode spurious correlations if the dataset is biased toward particular synthetic routes or spectral measurement conditions [12]. For chiral sensing applications, the representation must preserve information about multiple stereocenters, axial chirality, and supramolecular helicity. It must also be sensitive to the difference between enantiomers, which are often treated equivalently by conventional scalar molecular descriptors. This is a nontrivial demand because enantiomers have identical scalar graph structures in the absence of stereochemical annotations.

Supramolecular analytical chemistry provides an important source of experimental constraints for data representation and model validation. Optical sensing platforms based on metal-organic frameworks and related porous materials demonstrate how host-guest interactions can be transduced into measurable signals [18]. Supramolecular analytical methods also provide libraries of binding constants, selectivity factors, and chiroptical responses that can serve as training targets [19]. Stereodynamic probes illustrate how a chiral receptor can change its optical output upon guest binding, and their behavior depends on conformational dynamics that are challenging to capture with static representations [20]. Materials informatics frameworks that integrate these diverse data types can improve the generalizability of learned models, but they require careful alignment of experimental protocols and metadata [21]. The infrastructural lesson is that representation quality cannot compensate for poorly curated data, and that the most effective model architectures are those designed with an explicit awareness of how experimental measurements were generated.

4. Learning Paradigms and Model Governance

The selection of a learning paradigm for chiral supramolecular design involves a balance among supervised, unsupervised, generative, and reinforcement-based approaches. Supervised learning is suitable when labeled training data exist for properties such as binding affinity or chiral amplification. Unsupervised and generative models are valuable for exploring chemical space when labels are scarce, but their outputs require downstream validation. Reinforcement learning and active learning can be used to guide experimental campaigns by selecting the next most informative synthesis or spectroscopic measurement. In all cases, the model must be embedded within a governance structure that defines performance criteria, uncertainty thresholds, and intervention protocols. Best practices in machine learning for chemistry emphasize the importance of rigorous data splitting, uncertainty quantification, and benchmark transparency [22]. Without such governance, a model may produce plausible molecular structures that fail to assemble into the intended chiral organization or that cannot be synthesized under practical constraints.

Model governance also involves the management of model cards, documentation, and audit trails. The concept of model cards, originally developed in the broader machine learning community, is directly relevant to materials discovery workflows. A model card for a chiral supramolecular predictor would document the intended use, training data distribution, performance limitations, and known failure modes [23]. This documentation is especially

important when predictions are used to prioritize expensive synthesis or to decide between competing sensor designs. Because supramolecular systems are dynamic and environment-sensitive, a model that performs well for one solvent system may fail in another. Governance frameworks must therefore require explicit reporting of the experimental conditions under which training data were collected and the conditions under which the model is expected to be valid.

Another governance concern is the feedback loop between model predictions and experimental data generation. If model predictions are used to select which experiments to perform, and the resulting data are then used to retrain the model, the training distribution can become progressively biased. This is a form of distributional shift that is well known in active learning and adaptive experimentation. In chiral supramolecular design, such bias may cause the model to overrepresent easily synthesized monomer classes or readily measurable chiroptical signals while neglecting difficult but scientifically important regions of chemical space. Addressing this problem requires institutional mechanisms for data sharing, negative result reporting, and periodic independent audits. The governance challenge is therefore not only technical but also organizational. Research groups, funding agencies, and publishing venues need to coordinate standards for data deposition and model evaluation so that machine learning-assisted materials design remains cumulative and reproducible.

5. Structural Validation and Chiroptical Sensing Integration

Validation is the bridge between computational prediction and sensing application. In covalent molecular design, validation often focuses on whether a predicted structure can be synthesized and whether its calculated property matches experiment. In chiral supramolecular systems, validation must additionally address whether the intended supramolecular organization is actually formed and whether the chiroptical readout reflects the desired molecular recognition event. Circular dichroism spectroscopy, vibrational circular dichroism, and circularly polarized luminescence can reveal the presence and handedness of supramolecular assemblies, but these methods are sensitive to sample preparation and may not directly report on binding selectivity. The integration of chiroptical sensing into the design loop therefore requires multiple levels of validation: structural validation of the assembly, functional validation of guest recognition, and operational validation of the sensing device.

A central structural trade-off is between signal intensity and selectivity. Strong chiroptical signals can arise from large supramolecular aggregates, but such aggregates may bind guests nonspecifically. Smaller, well-defined receptor pockets may offer higher selectivity but produce weaker optical responses. Machine learning can assist in navigating this trade-off by predicting both chiroptical response and binding behavior across a library of candidate assemblies. However, the model must be trained on data that separate the contributions of chiral organization, guest binding, and environmental interference. The experience of supramolecular analytical chemistry shows that optical sensing is most effective when the receptor is designed to couple a specific binding event to a large change in molecular conformation or electronic structure [19]. The use of stereodynamic probes demonstrates that dynamic conformational changes can amplify chiroptical signals upon guest binding [20]. These design principles should be encoded as interpretable features or as constraints in the machine learning workflow.

The deployment of chiral supramolecular sensors introduces additional robustness requirements. Laboratory demonstrations often operate under carefully controlled conditions,

whereas real-world sensing occurs in complex matrices with fluctuating temperature, humidity, and interferents. A sensor architecture based on a single optimized material may be fragile outside its designed operating envelope. A system-level alternative is to deploy an array of receptors with partially overlapping selectivities and to use machine learning to decode the resulting response patterns. This array-based approach is common in electronic nose and tongue technologies, and it has conceptual parallels in supramolecular sensing. It shifts the design burden from achieving perfect selectivity in a single material to achieving reproducible differential responses across a receptor library. This strategy improves robustness because the failure of one receptor does not necessarily invalidate the array-level classification. It also enables the use of simpler chiral materials that might be insufficiently selective on their own but useful in combination. The trade-off is that array deployment requires more complex data processing and more extensive calibration, which must be addressed through careful system design.

6. Deployment, Robustness, and Sustainability

Deployment of machine learning-assisted chiral supramolecular sensors requires attention to the full lifecycle of the sensing system. This includes material synthesis, sensor fabrication, calibration, field operation, maintenance, and eventual disposal or regeneration. Sustainability considerations arise at multiple levels. The computational training of large generative models can carry significant energy costs, particularly when coupled with extensive quantum mechanical calculations. Experimental synthesis of candidate materials consumes solvents, reagents, and instrumental time. A design workflow that maximizes predictive accuracy without considering experimental cost may be scientifically productive but environmentally burdensome. System-level optimization should therefore include resource constraints, favoring materials that are synthesizable from readily available precursors, reusable across multiple sensing cycles, and stable under ambient conditions. Digital twin concepts from industrial production can inform this integration by providing a virtual representation of the material, the sensing device, and its operational environment [25]. In such a framework, model predictions are continuously updated with operational data, enabling predictive maintenance and adaptive recalibration.

Robustness in chiral sensing is closely linked to the distribution of training data and the treatment of uncertainty. A model trained on high-purity enantiomer samples may fail when confronted with racemic mixtures, structurally similar interferents, or degraded sensor surfaces. Robust deployment therefore requires adversarial and stress testing under realistic conditions. The machine learning component should report calibrated uncertainty estimates, and the system should be designed to request human intervention or additional measurements when confidence is low. This is particularly important in applications involving pharmaceutical quality control, environmental monitoring, or clinical diagnostics, where erroneous recognition events can have significant consequences. The governance structures described earlier must be extended to deployment so that model limitations are communicated to end users and regulators. Fairness concerns also emerge at this stage. If a sensor system is validated predominantly on datasets drawn from a narrow set of chemical environments or manufacturing conditions, its performance may be systematically worse for underrepresented analytes or operational settings. Broader discussions of representational harm and dataset bias in machine learning are directly relevant to materials sensing and should be integrated into validation protocols [24].

The policy implications of machine learning-assisted materials design extend beyond individual technical performance. Autonomous discovery platforms that combine generative models, automated synthesis, and high-throughput characterization may accelerate innovation, but they also raise questions about accountability, intellectual property, and safety. When a model proposes a candidate material, responsibility for verifying its properties and ensuring its safe use remains distributed across researchers, institutions, and regulatory bodies. Clear documentation standards and shared data repositories can reduce duplication and support independent verification. Publicly funded research infrastructures should consider open data mandates for materials informatics, while respecting legitimate commercial interests. Policy instruments may also be needed to ensure that the benefits of accelerated materials discovery are distributed across different sectors and regions, rather than concentrating in organizations with access to the largest computational and experimental resources. The alignment of machine learning-assisted design with sustainability goals will depend on such institutional arrangements as much as on algorithmic innovation.

7. Fairness, Policy, and Institutional Implications

A socio-technical perspective on chiral supramolecular design recognizes that technical systems are embedded within institutional environments. The selection of which materials to optimize, which sensor applications to pursue, and which performance metrics to prioritize is not a purely technical decision. It reflects funding priorities, regulatory demands, and societal needs. For example, enantioselective sensors are highly relevant to pharmaceutical manufacturing, food safety, and environmental monitoring, but these application domains have different tolerance thresholds and evidentiary standards. A machine learning workflow optimized for discovery throughput may not be well suited to regulatory contexts that require mechanistic interpretability and long-term reproducibility. Institutional frameworks must therefore differentiate between exploratory research models and validated decision-support tools.

Fairness in materials informatics is often discussed in terms of data coverage across chemical space. Just as algorithmic fairness in social systems requires attention to underrepresented groups, materials models require attention to underrepresented regions of the chemical and application space. If training data are dominated by well-studied metal complexes or easily synthesized aromatic systems, the resulting models may systematically underperform for chiral supramolecular assemblies based on less common scaffolds, metal centers, or solvent environments. This can reinforce existing research concentration and delay the discovery of materials for emerging sensing challenges. Mitigating this bias requires deliberate sampling strategies, open publication of negative results, and funding for data curation in underexplored domains. It also requires model developers to document the intended scope of use and the populations of chemical systems on which the model has been validated. Such documentation practices connect materials design to broader governance conversations about algorithmic accountability [23][24].

The long-term institutional challenge is to create a learning ecosystem in which computational predictions, experimental observations, and deployment failures are continuously integrated. This requires more than a shared database; it requires shared semantics, interoperable metadata, and incentives for data contribution. Academic credit systems often reward novel synthesis or high-impact predictions rather than careful negative results or infrastructure maintenance. To support machine learning-assisted chiral supramolecular design at scale, funding agencies and universities may need to recognize data curation, software maintenance,

and benchmark development as first-class scholarly contributions. Industrial and national laboratory partners can contribute by developing standardized testbeds for sensor validation and by sharing operational data under appropriate confidentiality frameworks. International coordination is also important because chiral sensing has implications for pharmaceutical supply chains and environmental regulation that cross national boundaries.

The governance of machine learning in materials discovery should be adaptive. Unlike conventional chemical safety regulation, which operates on relatively stable categories of substances, machine learning workflows can generate novel candidate structures at a rapid pace. This does not necessarily require entirely new regulatory paradigms, but it does require a shift toward process-oriented oversight. Regulatory agencies may need to evaluate not only individual materials but also the validation procedures, data provenance, and uncertainty reporting associated with a computational design platform. Process-oriented regulation can accommodate the dynamic nature of supramolecular design while preserving accountability. It also aligns with established quality management concepts in manufacturing and with emerging best practices for model reporting [22][23]. By embedding such governance mechanisms from the outset, the research community can build trust in machine learning-assisted materials design and reduce the risk of overclaiming or premature deployment.

8. Conclusion

Machine learning-assisted design of chiral supramolecular materials represents a promising convergence of computational chemistry, materials science, and analytical sensing. Yet its success cannot be reduced to the predictive accuracy of a single model. The design problem is inherently systemic, involving trade-offs between molecular detail and data scale, between model expressiveness and interpretability, and between laboratory optimization and real-world robustness. Chiral supramolecular systems are dynamic, environment-sensitive, and stereochemically complex, which means that data representation, experimental validation, and chiroptical sensing integration must be treated as coequal components of the research infrastructure. The layered architecture proposed in this paper highlights the need for bidirectional information flow between theoretical chemistry, machine learning, synthesis, and sensor deployment. It also emphasizes that infrastructure and governance are not peripheral concerns but central determinants of long-term productivity and safety.

The analysis has shown that data fragmentation, biased sampling, and weak model documentation are significant barriers to reliable deployment. Addressing these barriers requires institutional coordination across academic, industrial, and regulatory sectors. Model cards, open data standards, negative result reporting, and process-oriented oversight can improve accountability without stifling innovation. Sustainability considerations must also be integrated into the design loop, from the energy costs of computational screening to the environmental footprint of synthesis and sensor fabrication. Fairness concerns in materials informatics extend to chemical space coverage and application-domain representation, suggesting that the benefits of accelerated discovery should be deliberately broadened. Digital twin concepts and adaptive calibration strategies can improve operational robustness by linking laboratory models to field performance.

Looking forward, the most consequential advances will likely arise not from isolated algorithm improvements but from integrated systems that combine generative molecular design, chirality-sensitive validation, and adaptive sensing deployment. Such systems will require new forms of collaboration among computational scientists, supramolecular chemists, device engineers, data curators, and policy analysts. The required theoretical understanding of

how metal ions and other regulators control supramolecular chirality is an example of the mechanistic insight that must be embedded within data-driven workflows [17]. As these systems mature, they can support selective molecular recognition and sensing applications in pharmaceuticals, environmental monitoring, and beyond. The central challenge is to build not only better models but also better institutional arrangements that make those models trustworthy, reproducible, and responsive to societal needs.

References

1. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

2. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O., & Walsh, A. (2018). Machine learning for molecular and materials science. Nature, 559(7715), 547–555. https://doi.org/10.1038/s41586-018-0337-2

3. Chen, C., Ye, W., Zuo, Y., Zheng, C., & Ong, S. P. (2019). Graph networks as a universal machine learning framework for molecules and crystals. Chemistry of Materials, 31(9), 3564–3572. https://doi.org/10.1021/acs.chemmater.9b01294

4. Sturluson, A., Huynh, M. T., York, A. H. P., & Simon, C. M. (2019). Eigencages: Learning a latent space of porous cage molecules. ACS Central Science, 5(10), 1663–1676. https://doi.org/10.1021/acscentsci.9b00642

5. Le, T. C., & Winkler, D. A. (2016). Discovery and optimization of materials using evolutionary approaches. Chemical Reviews, 116(10), 6107–6132. https://doi.org/10.1021/acs.chemrev.5b00691

6. Ma, W., Xu, L., de Moura, A. F., Wu, X., Kuang, H., Xu, C., & Kotov, N. A. (2017). Chiral inorganic nanostructures. Chemical Reviews, 117(12), 8041–8093. https://doi.org/10.1021/acs.chemrev.6b00755

7. Liu, M., Zhang, L., & Wang, T. (2015). Supramolecular chirality in self-assembled systems. Chemical Reviews, 115(15), 7304–7397. https://doi.org/10.1021/cr500671p

8. Hembury, G. A., Borovkov, V. V., & Inoue, Y. (2008). Chirality-sensing supramolecular systems. Chemical Reviews, 108(1), 1–73. https://doi.org/10.1021/cr050005k

9. Yan, X., Wang, F., Zheng, B., & Huang, F. (2012). Stimuli-responsive supramolecular polymeric materials. Chemical Society Reviews, 41(18), 6042–6065. https://doi.org/10.1039/c2cs35091b

10. Amabilino, D. B., Smith, D. K., & Steed, J. W. (2017). Supramolecular materials. Chemical Society Reviews, 46(9), 2404–2420. https://doi.org/10.1039/c7cs00163k

11. Schneider, H.-J. (2009). Binding mechanisms in supramolecular complexes. Angewandte Chemie International Edition, 48(22), 3924–3977. https://doi.org/10.1002/anie.200802947

12. Gryn'ova, G., Lin, K. H., & Corminboeuf, C. (2018). Read between the molecules: Computational insights into organic semiconductors. Journal of the American Chemical Society, 140(48), 16370–16386. https://doi.org/10.1021/jacs.8b07985

13. 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

14. Pyzer-Knapp, E. O., Suh, C., Gómez-Bombarelli, R., Aguilera-Iparraguirre, J., & Aspuru-Guzik, A. (2015). What is high-throughput virtual screening? A perspective from organic materials discovery. Annual Review of Materials Research, 45, 195–216. https://doi.org/10.1146/annurev-matsci-070214-020823

15. Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., Li, B., Madabhushi, A., Shah, P., Spitzer, M., & Zhao, S. (2019). Applications of machine learning in drug discovery and development. Nature Reviews Drug Discovery, 18(6), 463–477. https://doi.org/10.1038/s41573-019-0024-5

16. Jablonka, K. M., Ongari, D., Moosavi, S. M., & Smit, B. (2020). Big-data science in porous materials: Materials genomics and machine learning. Chemical Reviews, 120(16), 8066–8129. https://doi.org/10.1021/acs.chemrev.0c00004

17. Cai, Yiteng, et al. "Theoretical Chemistry Facilitated Understanding of Supramolecular Chirality Regulation by Metal Ions." ACS Materials Letters 8.2 (2026): 621-627.

18. Lustig, W. P., Mukherjee, S., Rudd, N. D., Desai, A. V., Li, J., & Ghosh, S. K. (2017). Metal–organic frameworks: Functional luminescent and photonic materials for sensing applications. Chemical Society Reviews, 46(11), 3242–3285. https://doi.org/10.1039/C6CS00930A

19. You, L., Zha, D., & Anslyn, E. V. (2015). Recent advances in supramolecular analytical chemistry using optical sensing. Chemical Reviews, 115(15), 7840–7892. https://doi.org/10.1021/cr5005524

20. Wolf, C., & Bentley, K. W. (2013). Chirality sensing using stereodynamic probes with distinct chiroptical responses. Chemical Society Reviews, 42(12), 5408–5424. https://doi.org/10.1039/c3cs35498a

21. Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A., & Kim, C. (2017). Machine learning in materials informatics: Recent applications and prospects. npj Computational Materials, 3, 54. https://doi.org/10.1038/s41524-017-0056-5

22. Artrith, N., Butler, K. T., Coudert, F.-X., Han, S., Isayev, O., Jain, A., & Walsh, A. (2021). Best practices in machine learning for chemistry. Nature Chemistry, 13(6), 505–508. https://doi.org/10.1038/s41557-021-00716-z

23. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 220–229). ACM. https://doi.org/10.1145/3287560.3287596

24. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610–623). ACM. https://doi.org/10.1145/3442188.3445922

25. Uhlemann, T. H.-J., Lehmann, C., & Steinhilper, R. (2017). The digital twin: Realizing the cyber-physical production system for Industry 4.0. Procedia CIRP, 61, 335–340. https://doi.org/10.1016/j.procir.2016.11.152