

An MIT Exploration of Generative AI • From Novel Chemicals to Opera

Closing the Execution Gap in Generative AI for Chemicals and Materials: Freeways or Safeguards

**Akshay Subramanian¹ Wenhao Gao² Regina Barzilay^{3,4}
Jeffrey C. Grossman¹ Tommi Jaakkola^{3,4} Stefanie Jegelka^{3,4,5}
Mingda Li⁶ Ju Li⁶ Wojciech Matusik³ Elsa Olivetti¹ Connor W. Coley^{2,3}
Rafael Gomez-Bombarelli¹**

¹Department of Materials Science and Engineering, MIT, Cambridge, Massachusetts 02139, USA,

²Department of Chemical Engineering, MIT, Cambridge, Massachusetts 02139, USA,

³Department of Electrical Engineering and Computer Science, MIT, Cambridge, Massachusetts 02139, USA,

⁴Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, Massachusetts 02139, USA,

⁵Department of Computer Science, TU Munich, 85748 Garching, Germany,

⁶Department of Nuclear Science and Engineering, MIT, Cambridge, Massachusetts 02139, USA

MIT

Published on: Mar 27, 2024

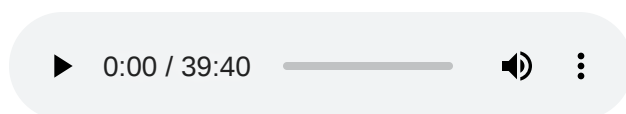
DOI: <https://doi.org/10.21428/e4baedd9.92e511e3>

License: [Creative Commons Attribution-NonCommercial 4.0 International License \(CC-BY-NC 4.0\)](https://creativecommons.org/licenses/by-nc/4.0/)

ABSTRACT

Designing chemicals and materials is a winner-takes-all game in a chemical space. Out of a nearly endless list of candidate chemicals, only a tiny subset can practically be made and tested for a given task. Of those, a mere handful will advance to the clinic or the market because of regulatory and economic factors, like Food and Drug Administration approval or Capital Expenditure for manufacturing. This attrition is slow, expensive, and high stakes; committing to the wrong chemical can ruin a clinical trial or doom a consumer device. Generative AI has undoubtedly broadened and accelerated the early stages of chemical design. However, real-world success takes place further downstream, where the impact of AI has been limited so far. Here, we identify the nature of this ‘execution gap’ and analyze its technical and social sources with a focus on the domain-specific nuances that separate chemistry and materials from other applications of AI. We explore strategies to address the gap and include directions for algorithmic development, strategies for efficient creation and sharing of data, and maximization of feedback between early discovery and downstream validation. Like any powerful technology, generative AI in chemistry brings ethical concerns and dual-use risks. In our analysis, these are not as critical or immediate as they are in other applications of AI. Precisely because of the execution gap, AI has so far not meaningfully lowered the barrier for bad actors to produce chemicals known to be harmful and nor has AI creativity practically broadened the risk spectrum in terms of the types and severity of potential misuse.

Keywords: drug discovery; autonomous laboratories; sustainability; materials science



Listen to this article

1. Introduction

The use of generative AI in chemistry and materials has closely tracked the evolution of the field and at times guided algorithmic innovations ([Figure 1](#)).[\[1\],\[2\],\[3\]](#) In the essentially infinite chemical space, generative AI offers the promise of inverse design by automatically producing winning designs instead of iterating over candidate chemicals by trial and error. This idea has been demonstrated: AI algorithms have expedited the discovery of small molecule drugs;[\[4\],\[5\]](#) protein structures created by generative AI have been experimentally verified;[\[6\],\[7\]](#) large language models, trained on scientific literature, have been used to design and plan the synthetic pathways of metal organic frameworks[\[8\],\[9\]](#) and to orchestrate data access and robotic experimentation;[\[10\]](#) and the concept of AI-guided closed-loop discovery, integrated with autonomous experimentation, has been demonstrated in multiple proofs of concept ([Figure 2](#)).[\[11\],\[12\],\[13\],\[14\]](#) The successes achieved beg the question: Is it possible to conjure up a blockbuster drug or a high-temperature superconductor from a well-stated user prompt, similar to consumer-facing generative models for text or

images? Such capabilities could alter the scientific discovery landscape by lowering knowledge and cost barriers to design new chemicals. Naturally, concerns arise that misuse of such technology may enable the creation of illicit or toxic substances, warranting new safeguards and regulations.[\[15\]](#),[\[16\]](#)

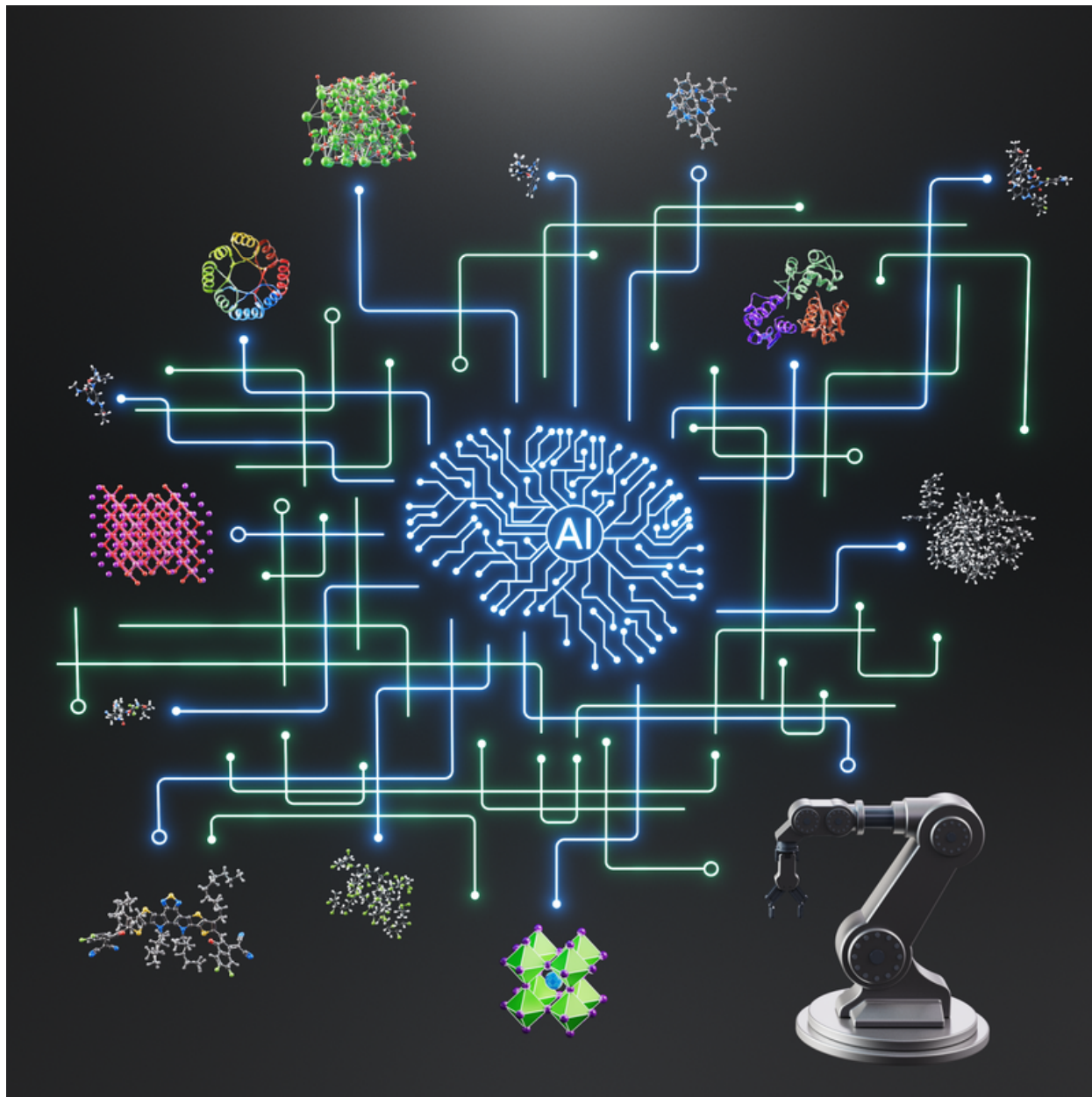


Figure 1

Schematic illustration of generative AI and automation for chemistry and materials science.

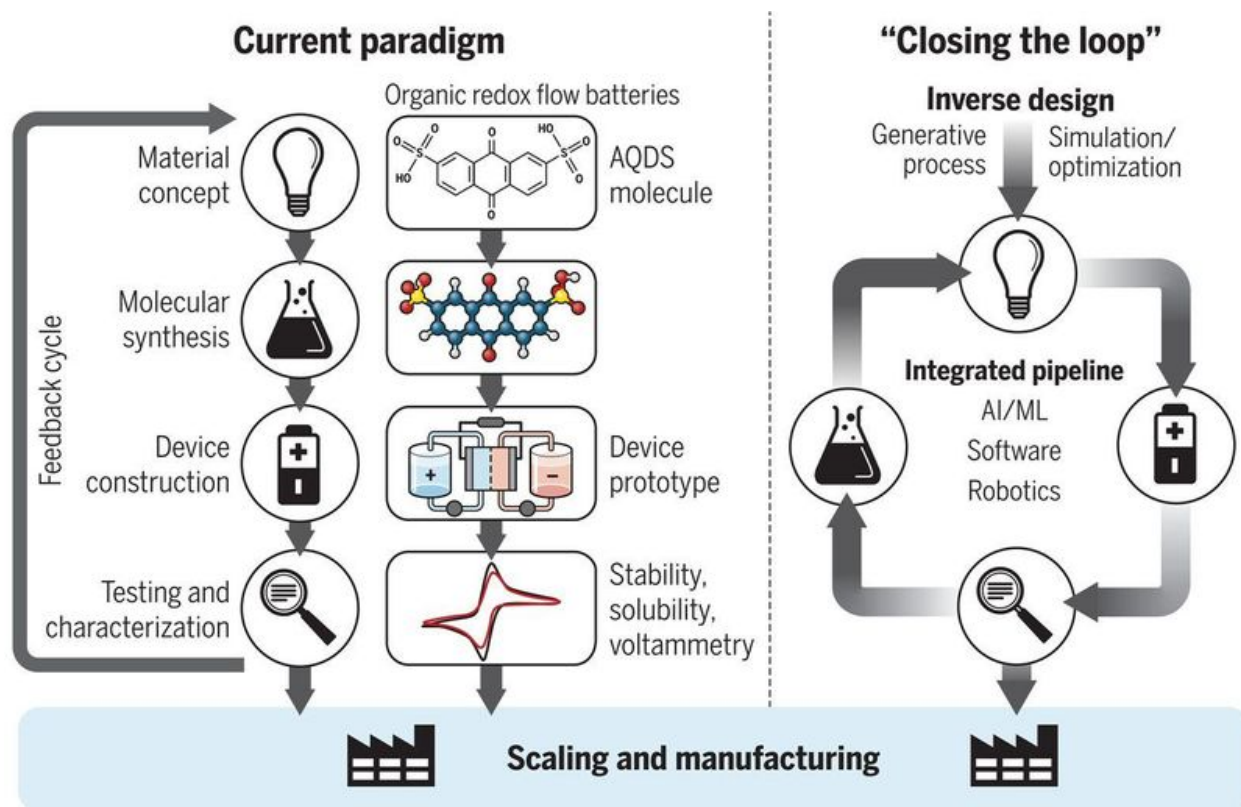


Figure 2

Comparison between traditional and AI-guided design workflows. [Reproduced from reference 1]

In reality, generative AI is not yet an established success—or risk—in chemistry and materials science. Nearly a decade has elapsed since the first deep learning generative models for chemistry were published; reported victories are increasing but remain preliminary. While promising, innovations that can be uniquely attributed to AI have not been carried forward to the clinic or scaled up as consumer products and are within the reach of just a few specialized laboratories.

Some existing barriers are likely addressable through algorithmic innovation that would generalize applicability. Unlike vision or natural language models, machine learning (ML) models in chemistry face challenges when operating beyond their training domain. Trustworthy extrapolation, essential for scientific discovery and innovative designs, is lacking. This limitation necessitates expensive experimental validation of potentially inaccurate ideas.

More critically, a practical execution gap exists between generative AI and the physical world. In chemistry and materials, outcomes must be embodied, evaluated for function and safety, scaled up, and commercialized. Despite the rapid generation of ideas, evaluating them is costly, involving small-scale laboratory testing, clinical trials, and slow to set up pilot plants, which produce sparse data. Current AI struggles with these downstream stages and lacks the ability to learn from them. The definition of ease of synthesis and scalability

are themselves application dependent and may involve precursor cost, supply chain availability, adaptability of industrial infrastructure, reproducibility, yield, etc. This friction between computation and translation is not uniform across applications because of algorithmic reasons, availability of open data, speed and throughput of experimentation, and economic drivers of innovation. Social factors are relevant too. Scientists may be demotivated to carry out the designs of an AI that could be inscrutable, error prone, overconfident, or unoriginal. Likewise, there are instances of disagreement between the AI community and traditional subject matter experts on the size and nature of this gap and on how to measure AI success.

The reality of this execution gap must also be considered when addressing dual-use concerns. In theory, the difference between minimizing toxicity for food coloring and maximizing it for a chemical weapon is merely a negative sign, but most generative AI approaches to molecular design are fraught with mispredictions and critically bottlenecked by time-consuming empirical evaluation. The community lacks a standardized set of ethics and risk management rules, unlike human and animal experiments or the emerging standards in the AI community. The risks of purely digital innovations may be overblown, disregarding the difficulty of actually making chemicals, and the corresponding risk management strategies may be inordinately cautious, blocking the release of algorithms, training data, or even user inputs. Since embodiment is a technological chokepoint, it is also the natural place for guardrails: controlling access to precursors or laboratories is both easier and more meaningful than keeping technical programming decisions or training data secret.

Distributed efforts, including algorithmic innovation, open data initiatives, and collaboration between tech, chemical industry, and academia can help deliver generative AI's potential in chemistry and materials for healthcare and sustainability. These efforts can bridge the gap between digital discovery and physical productization while maintaining barriers against misuse. Although strides are being made, the pace of these efforts currently lags behind advances in generative AI.

2. Gaps in Using Generative AI for Chemistry and Materials

2.1. Learning Generalizable Representations

Chemistry and materials discoveries involve identifying physical matter with unprecedented properties, an extrapolation from known to unknown, which is particularly challenging because it takes place in a regime where training data is scarce. Furthermore, to invent *useful* chemicals and not just *random* chemicals, generative models must be coupled with discriminative models (i.e., property prediction models) trained on expensive labels from laboratory experiments or accurate simulations. Generative and discriminative models struggle to generalize in out-of-domain (OOD) regions. For example, a generator might lack creativity to propose novel chemistries beyond its training data, like suggesting a carbon fiber–reinforced plastic for an aviation airframe, which a model trained on metal alloys might overlook. Additionally, property predictors trained on the same data may assign spurious fitness labels to OOD suggestions. Hallucinations, in which models are creative but factually wrong, are common in generative models, especially in this extrapolative

regime. Pretraining models on a large, broad dataset is a common technique in generative AI to create foundational models that can be efficiently fine-tuned on downstream tasks. Pretraining and foundational models have seen great success in images and text and have also been relatively successful in the space of protein design but—despite many efforts—not as much for molecules and even less so for solid crystalline materials, which have only recently started to be addressed.[\[17\]](#),[\[18\]](#),[\[19\]](#) Unlike small molecules, the properties of solid materials cannot be captured only from the structure of a few atoms and depend on processing as much as on composition, and for instance, the design of new catalysts requires algorithms capable of going beyond small atomic structures.[\[20\]](#),[\[21\]](#) This broad challenge in generalizability is one reason why closed-loop, iterative, experimental validation is so vital in a typical workflow and why developing meaningful and comprehensive training datasets has been such an important focus.

2.2. Generating Training and Validation Data

Laboratory experiments are slow, expensive, and poorly scalable, which usually make them the rate-determining step in an ML-driven discovery pipeline. The significance of this problem has been reflected in several efforts to accelerate the experimental validation process with autonomous laboratories in the past half decade.[\[11\]](#),[\[13\]](#),[\[22\]](#),[\[23\]](#),[\[24\]](#),[\[25\]](#),[\[26\]](#),[\[27\]](#) They have, however, experienced some unique practical challenges.[\[28\]](#),[\[29\]](#) For instance, the products that these robotic platforms can synthesize are limited by the set of prespecified precursor chemicals and implemented physical operations. Narrower chemical and operational ranges allow for more efficient automation, as evidenced by the robustness and commercial availability of DNA, RNA, and peptide synthesizers. With a greater breadth of the molecular products a robotic platform is able to make comes greater cost and complexity. Producing a new molecule, material, or device often requires a complex sequence of physical operations beyond mixing and heating, not the least of which is designing purification and isolation strategies to ensure that the properties one measures correspond to the molecule one intended to make. In addition, current off-the-shelf robotic arms lack the dexterity and coordination of human hands, which is crucial for tasks like the dosing of solid reagents with diverse textures.[\[30\]](#) While the initial examples of automated labs have been successful proofs of concept for accelerated synthesis and closed-loop design workflows, they also highlighted the challenges: the time and capital investment in developing and installing these platforms is large, so the averaged time and cost per unit experiment are not necessarily lower than traditional experimentation. It is also likely that experimental validation will still remain the rate-determining step, even with the development of general purpose robotic labs in the future, placing high importance on the predictive quality of generative models.

2.3. Data Quantity and Quality

The performance of both generative and discriminative models is bounded by the quality of the training dataset. Labeled data in the chemical sciences is obtained from experiments that are susceptible to both instrumental and human errors. For instance, the signal-to-noise ratio in chemical and biochemical experiments can be small, particularly across replicated biological experiments, which makes the direct application of

machine learning to results from human experiments challenging.[\[31\]](#) The quantity of the data is also an important consideration since the state-of-the-art models employing end-to-end learning typically demand large training datasets. Autonomous experiments could potentially enhance both quality and quantity, but they still face the challenges mentioned above. Up to now, ‘real’ success stories such as approved drugs or commercialized materials, whether aided by AI or not, are few and provide sparse reward signals to gauge the superiority of one discovery workflow over another.

Data bias is another concern, as it leads to biased predictions. This includes the more obvious issue of bias, that is, erroneous labels, but also the distribution mismatch between the training data and the target chemical space. In molecular design, it is unlikely for a generative model to generate candidates containing chemical scaffolds (or crystal structures in inorganics for instance) that are not well represented in the training dataset, even if they might lead to better properties. Likewise, training models on data from scientific literature[\[32\]](#) inherits reporting biases, posing a risk of systematically overoptimistic models.

Physics-based simulations can sometimes be a cheaper—but lower fidelity—alternative to experimental measurements, so it is common to use these synthetic labels to expand the coverage of models in chemistry. The choice of the size and time scale of simulations is an important factor. For instance, phenomena such as the degradation of metallic materials in natural environments depend on local microstructural features at a scale of micrometers to millimeters, which can only be observed with mesoscale simulations rather than atomistic simulations.[\[33\]](#) Datasets containing labels of multiple experimental fidelities, combining simulated and experimental data or various time and size scales, are and will continue to remain an integral part of chemical machine learning, which requires the development of algorithms able to exploit these dimensions of information.[\[34\],\[35\],\[36\],\[37\],\[38\]](#)

2.4. You Get—At Best—What You Ask For

On the algorithmic side, generative models must not only suggest novel candidates but also optimize them for desired properties. It has historically been difficult to clearly describe—in machine readable format—the functional requirements that accurately capture real-world complexities. For instance, efficacious drug molecules should not only demonstrate a high affinity towards specific targets but also exhibit selectivity against alternative targets, coupled with a favorable ‘drug-likeness’ profile, encompassing acceptable absorption, distribution, metabolism, excretion, and toxicity characteristics, and be synthetically accessible at the appropriate scale while meeting regulatory standards of purity, etc. We not only require quantitative estimates of all these properties but also need to understand the tradeoff between all these objectives during property optimization.

2.5. Success Is Hard (To Measure)

As generative AI in chemistry and materials transitions from a scientific novelty to a practical technology, different coexisting measures of success must be reconciled. Proof-of-principle instances of AI creativity or

autonomy may look trivial, or below standard, when compared with expert reasoning or traditional scientific rigor. The Automatic Chemical Design (ACD) levels, for instance, provide a means to classify the autonomy of agents based on the degree of contributions by chemists and machines in ideation and decision stages of the workflow.[\[39\]](#)

Recent instances such as the work on a deep learning method to identify potent DDR1 kinase inhibitors in *Nature Biotechnology*[\[5\]](#) and the work on an autonomous laboratory for materials synthesis in *Nature*[\[13\]](#) are examples of this difference in standards. In the former, the best candidate generated by the generative algorithm was structurally similar to ponatinib, a known DDR1 kinase inhibitor, posing the question of whether this was as ‘revolutionary’ as suggested by some media.[\[40\]](#),[\[41\]](#) In the latter, the phase identification and Rietveld refinement of X-ray diffraction patterns after synthesis was performed in an automated fashion with the usage of ML components, resulting in a poor quality of refinement as compared to the standard expected from human scientists, which might have led to inaccurate predictions of compositions.[\[42\]](#),[\[43\]](#) While methodological advances were quite significant in both cases, friction arose between holding the AI to its own standard versus the scientific standard in the given field.

A similar clash of standards sometimes arises around data and model availability. The gold standard for openness within the AI community is open sourcing of data and code,[\[44\]](#),[\[45\]](#) while publications in chemistry and materials are expected to comply with high standards of disclosure in methodology and result validation. Ideally, research articles at the interface would apply the superset of both expectations, but AI works in chemistry and materials are at risk of both underreporting experimental details compared with traditional works in the field and not making their code and data available (in a way that is usable). Releasing training data, code, and model checkpoints along with published papers should be a mandatory requirement in journals when the authors hold the intellectual property rights. In works coupling AI with robotic platforms, it is a reasonable choice to not make access to robotic hardware controls publicly accessible for safety concerns. However, not releasing accompanying software and datasets impedes scientific progress by blocking the community from accessing useful tools. Furthermore, it prevents due scientific scrutiny. Lastly, commercial interests in developing software or chemicals may be perfectly justified, but they should not interfere with reporting standards and must be accompanied with appropriate disclosures in publications. It is particularly dangerous for the community to mask for-profit or other spurious interests with poorly justified misuse concerns as a reason to underreport.

3. Paths Forward

Bridging the execution gap of generative models in chemical science includes tackling components arising from (1) the performance of generative algorithms and (2) the expense of experimental validation.

3.1. Algorithm Innovation

The unique challenges posed by scientific tasks demand algorithmic innovation in areas that may not be the focus for the development of general-purpose AI tools. Inverse chemical design models must couple generative AI with property prediction models to drive design to novel structures that optimize for properties of interest. While the first developments were seen in molecules [46], [47], periodic materials have gained traction more recently [48], [49]. Improving predictive quality and generalization is critical to maximize success rates and avoid endless data acquisition loops focused on repairing unreliable models. [50] Since there is some overlap in problems and solutions for both generative and discriminative models, we discuss both cases within a common framework.

Pretraining, transfer-learning, and foundation models. Datasets in chemical science suffer from sparse real-world successes and expensive labeling, which often translates to small training datasets, so designing optimized molecules with as little labeled data as possible is still an open question. [51] Improving the effect of pretraining and transfer-learning strategies in AI for chemistry is critical. Being able to train foundation models for chemistry that can be fine-tuned for specific tasks is a promising path to address the scarcity of quality labeled data. There have been promising attempts in the direction of transfer learning for chemistry recently, especially for property prediction tasks, but they are yet to reach the same level of effectiveness as language and vision models. [52], [53], [54] Maximizing the usage of existing publicly available datasets across domains will help drive such foundation models. Recent works suggest that generalized architectures for chemistry and materials follow favorable scaling laws and can continue to improve with larger datasets. [55], [56] Novel approaches bridging the gap between pretraining and transfer learning strategies and the intricacies of scientific tasks remains a pivotal focus for advancing the field.

Inductive bias. Incorporating known science-based, domain-specific priors into models could be another avenue for extracting value from small datasets. A recent successful trend is the use of equivariant models that incorporate spatial symmetry rigorously, such as the E3NN framework. [57] By having the internal state and the operation of the model respect strictly the symmetry constraints of their inputs and labels, equivariance improves data efficiency and physical faithfulness in some classes of prediction tasks. [58], [59], [60], [61], [62], [63], [64] It is, however, uncertain how effective such physical priors are in generative models, with some recent research showing equivalent generalization performance being achieved without domain-specific inductive biases. [3], [65] There is hence a need for further research to explore the coupling between physical priors, dataset sizes, and the task being performed with generative models.

Digitizing synthetic accessibility. A direct adoption of graph generative algorithms for designing molecules exhibits severe problems in synthetic accessibility, [66] which motivates considering synthetic accessibility during generation. [67], [68], [69], [70], [71] Different chemical entities—such as small molecules, proteins, RNA, and solid materials—present unique challenges in generative modeling. The formalisms of small molecules as graphs, proteins as sequences of canonical amino acids, and RNA as sequences of nucleotides do

not necessarily have straightforward analogs for solid materials with periodicity, certain defect structures, or well-defined microstructure that arise from the interplay of composition and processing. Consequently, the creation of models specifically tailored to each of these entities is vital.[\[49\]](#),[\[72\]](#),[\[73\]](#),[\[74\]](#) On the other hand, given that all chemical species adhere to universal physical principles and given that many applications demand simultaneous modeling of different chemical entities (e.g., small organic molecules and proteins in drug discovery or zeolites and organic structure-directing agents in catalysis), there is an expectation for models to generalize across diverse chemical spaces and multiple length scales, exemplified by initiatives like RosettaFoldAA[\[75\]](#) or so-called foundational machine learning potentials in materials.[\[56\]](#),[\[76\]](#),[\[77\]](#),[\[78\]](#) ML algorithms can also find use in accelerating experimental validation, such as by planning chemical synthesis, which is particularly promising when coupled with automated laboratories.[\[79\]](#),[\[80\]](#),[\[81\]](#)

Uncertainty quantification and active learning. To tackle experimental challenges in throughput and scalability, one possible algorithmic approach involves enhancing the efficiency of design and optimization workflows while easing the burden on validation components. Active learning (AL) strategies, utilizing Bayesian optimization for instance, can be employed to minimize costly queries/experiments, particularly in machine learning–accelerated synthesis workflows.[\[82\]](#),[\[83\]](#) Uncertainty estimates of machine learning models, that is, knowing when the model prediction is not reliable, are a vital part of the data acquisition strategies in AL. Existing approaches for uncertainty quantification, while varied, face limitations tied to specific datasets and tasks, hindering widespread adoption.[\[84\]](#) Developing universally applicable, well-calibrated uncertainty quantification techniques in chemistry and materials would enhance the efficiency of AL strategies.

3.2. Experimental Capabilities

A promising solution to accelerate the data generation and experimental validation is using robotic execution to achieve automated, and eventually autonomous, laboratories. Automation in chemical experiments is the first step towards autonomous experimentation, based on the modularization and scaling of basic experimental operations.[\[11\]](#),[\[13\]](#),[\[14\]](#),[\[24\]](#),[\[25\]](#),[\[26\]](#),[\[27\]](#) Many of these operations have commercialized solutions, such as liquid handling and plate or vial transfer, but certain key steps still lack effective commercial solutions, such as accurately handling powdered solid and viscid substances or very small liquid volumes. Integrating characterization to effectively ‘test and analyze’ and thus close the loop also challenges current workflows. There are both hardware and software opportunities. Characterization instrumentation is very diverse and typically costly, and only some techniques are amenable to physical integration alongside the automated synthesis equipment, and many others reside in physically separated facilities. This is particularly meaningful for solid materials, since their properties arise from complex multiscale interactions. Depending on the level of integration and physical proximity of the characterization equipment, samples can be transferred using fluidic techniques, stationary robotic arms, or mobile robotic arms.[\[85\]](#) On the software side, ML techniques offer the potential for automated analysis of characterization data.[\[86\]](#) The integration of heterogeneous characterization

data from multiple sources, such as X-ray crystallography, microscopy, nuclear magnetic resonance, or various spectroscopic techniques, is also an ongoing area of research.

Autonomous experimentation requires using AI planning to fully or partially replace detailed human-written instructions. This involves collaborative advancements in synthesis planning algorithms, algorithms for chemical hypothesis generation (molecular design), communication across multiple automated laboratories, and the ability to monitor experimental progress through multimodal sensing systems to verify experiment success. Some issues demand breakthroughs in fundamental science, such as analyzing chemical components in a mixture without the use of authentic reference standards at scale. Overall, the main challenges are often interdisciplinary engineering problems, requiring collaborative efforts from mechanical engineers, electrical engineers, chemical engineers, and others. Such challenges are not typically suited for resolution by individual academic teams. Government or private-backed user facilities, as they exist for materials characterization (beamlines) or biomedical research (contract research organizations), are a possible pathway to ameliorate the capital and operating expense of autonomous labs and maximize their impact.[\[30\]](#),[\[85\]](#),[\[87\]](#)

3.3. Data, Benchmarks, and Other AI Infrastructure

The development, training, evaluation, and deployment of AI algorithms still require additional components, which we refer to as infrastructure.[\[88\]](#) Among these, as mentioned above, data quality and quantity is a big issue. On the one hand, we need to accelerate data generation by promoting high-throughput experiments (see above). On the other hand, we need to encourage the production of more open data and reforms in data-sharing practices. Emphasizing the importance of consistent measurements in molecular and materials properties, it is crucial for government agencies or nonprofit organizations to invest in the development of open standard datasets. A prime example is the creation of large-scale, high-quality datasets similar to the National Institutes of Health's Tox21,[\[89\]](#) which has proven to be valuable for evaluating chemical predictive models.

Even when open data is available, locating and curating it and evaluating model performance in task-appropriate ways are nuanced and time-consuming tasks that require domain expertise.

To mitigate these issues, the creation of centralized, expert-maintained databases and benchmarks by authoritative institutions is crucial. For example, the Protein Data Bank (PDB),[\[90\]](#) a database of protein structures, which facilitated the establishment of Critical Assessment of Structure Prediction (CASP), a competition for benchmarking protein structure prediction models. Even before AI was a thinkable solution to CASP, the data availability enabled by PDB and the regular assessment of computational models enabled by CASP have been pivotal in the evolution and success of AI models and directly catalyzed the development of AlphaFold.[\[91\]](#) Initiatives such as the Therapeutic Data Commons[\[92\]](#) and the Open Reaction Database,[\[93\]](#) are also making strides in providing uniform access to open data in therapeutic science and organic reaction, respectively, marking progress towards resolving the highlighted challenges in data utilization and model evaluation. Moreover, it is important that publishers of scientific journals must support open data efforts, from

requiring digital supplementary information to unified standards for machine readable versions of figures and chemical compound names, so that they can readily be utilized as training data for models, thus simplifying the need for complicated literature and figure mining algorithms.[\[50\]](#),[\[94\]](#),[\[95\]](#),[\[96\]](#),[\[97\]](#)[\[98\]](#),[\[99\]](#),[\[100\]](#),[\[101\]](#),[\[102\]](#)

4. Ethical and Dual-Use Risks and Mitigation Strategies

The gap between generative model predictions and practical execution, observed in positive-use cases, also extends to malicious-use cases because of the same issues of predictive reliability, synthetic accessibility, and experimental production of novel chemicals. Recent work highlights concerns about generative models being used for malicious purposes, citing an example of designing toxic molecules with predicted LD50 values lower than nerve agent VX.[\[16\]](#) However, current generative models are far from reaching the potential to accurately design synthesizable compounds with specific property profiles, especially when compared to human experts.

Yet, as generative AI becomes an increasingly viable avenue for scientific discovery, its use in accelerating the design of novel harmful compounds or predicting synthetic routes for known harmful compounds could become more feasible. The inability of current models to reliably perform these tasks on basic molecules like aspirin [\[10\]](#) is not a viable safeguard in the long term.

4.1. Information Risk and the Broader Public

If, however, the objective shifts from instructing generative models to design compounds with novel properties to a more straightforward task, such as enhancing knowledge access about existing toxic chemicals, ML models and tools like conversational AI bots can pose significant risks on a shorter timeline. This was demonstrated in a recent incident in which a conversational meal-planner bot recommended a recipe for creating chlorine gas when guided to.[\[103\]](#) In situations in which the level of expertise of the end user and anticipated intent of usage are highly variable, it is crucial to balance the level of autonomy of the agent and provide due safeguards on what decisions are allowed and disallowed. This is a challenge for conversational AI in general (which can provide just as objectionable advice about dieting, medication, or investment) and must be addressed cohesively through alignment practices for generative AI, relying on ethics boards, legislative guidance, community engagement, and continuous oversight to implement strong safety nets.

4.2. Power-User Bad Actors

Existing safeguards in conversational AIs cannot consistently repel voluntary misuse by bad-actor power users in the form of prompt hacking and attacks, and this challenge carries over to chemical information. The level of risk that arises from exposing ‘forbidden’ chemical knowledge should be compared to the amount of information that would be available on the internet to bad actors with equal levels of motivation and expertise. As of today, actually making dangerous chemicals is harder than just knowing how to create them (which is harder than imagining what to create, as in the nerve agent example) because it requires access to specialized equipment and chemicals, on top of the combination of background technical knowledge and misuse-specific

information. It is foreseeable that AI could supplant some or all of the latter, making hardware and hands-on knowhow limitations the only blocking elements.

4.3. Execution Risk

There is a more unique scope for dangerous situations for AI that is applied further downstream. This is particularly important as experimental facilities gain more autonomy, and chances for candidates escaping close human inspection increases.^{[11],[13],[14],[24],[25],[26],[27]} In automated workflows, either fully autonomous or human-in-the-loop, it is important that questions on who proposes experiments, who approves/selects experiments, and who executes experiments are carefully considered within, for instance, the ACD levels framework so that responsibilities of the various parties involved are clearly delineated. Establishing guidelines and safeguards in this regard could be useful in preventing both intentional and unintentional misuse of autonomous platforms. This level of risk is more reminiscent of critical decisions like autonomous weapons systems, which—for clear reasons—are receiving strong attention. AI agents capable of *making* chemicals must receive very rigorous oversight, and the community needs to first develop the institutions to decide the nature of these limitations and then to enforce them.

The interface between ML model predictions and robotic hardware is a critical nexus for implementing checks. Methods may involve rule-based filters to ensure adherence to safety guidelines or advanced approaches with human feedback for thorough screening. The latter of course comes at the expense of lower overall throughput, so the level of regulations can be made to be commensurate with the level of risk and autonomy of the agent.

In addressing broader safety and misuse concerns of generative AI, President Biden's recent executive order, issued on October 30, 2023, mandates AI companies to disclose red-teaming exercise results and large-scale model training to the government. It also tasks the Department of Energy with investigating AI's potential role in cyberattacks and biological and chemical weapons development.^[104] Yet, the need still remains for policies tailored to the aforementioned, more specific questions that chemistry and materials science pose.

5. Conclusions

Generative AI offers both promise and challenges in reshaping molecule and materials discovery. The journey from concept to application faces computational limitations, ethical concerns, and a practical execution gap. As AI advances towards expert-free chemical realization, ethical considerations become crucial, requiring a balanced approach to prevent misuse and ensure responsible innovation. Collaboration across domains, open data initiatives, algorithmic advancements, and the fusion of computer science with domain expertise are vital for advancing scientific ML. This paper emphasizes the need for a balanced approach, leveraging generative AI's potential while addressing ethical and safety considerations in this evolving field.

References

1. Sanchez-Lengeling, Benjamin, and Alán Aspuru-Guzik. 2018. “Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering.” *Science* 361 (6400): 360–65. <https://doi.org/10.1126/science.aat2663>. ↵
2. Wang, Hanchen, Tianfan Fu, Yuanqi Du, Wenhao Gao, Kexin Huang, Ziming Liu, Payal Chandak, et al. “Scientific Discovery in the Age of Artificial Intelligence.” *Nature* 620, no. 7972 (2023):47–60. <https://doi.org/10.1038/s41586-023-06221-2>. ↵
3. Yang, Mengjiao, KwangHwan Cho, Amil Merchant, Pieter Abbeel, Dale Schuurmans, Igor Mordatch, and Ekin Dogus Cubuk. “Scalable Diffusion for Materials Generation.” *arXiv* (2023). <http://arxiv.org/abs/2311.09235>. ↵
4. Stokes, Jonathan M., Kevin Yang, Kyle Swanson, Wengong Jin, Andres Cubillos-Ruiz, Nina M. Donghia, Craig R. MacNair, et al. “A Deep Learning Approach to Antibiotic Discovery.” *Cell* 180, no. 4 (2020):688-702.e13. <https://doi.org/10.1016/j.cell.2020.01.021>. ↵
5. Zhavoronkov, Alex, Yan A. Ivanenkov, Alex Aliper, Mark S. Veselov, Vladimir A. Aladinskiy, Anastasiya V. Aladinskaya, Victor A. Terentiev, et al. “Deep Learning Enables Rapid Identification of Potent DDR1 Kinase Inhibitors.” *Nature Biotechnology* 37, no. 9 (2019):1038–40. <https://doi.org/10.1038/s41587-019-0224-x>. ↵
6. Dauparas, J., I. Anishchenko, N. Bennett, H. Bai, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, et al. “Robust deep learning-based protein sequence design using ProteinMPNN.” *Science* 378, no. 6615 (October 2022):49–56. <https://doi.org/10.1126/science.add2187>. ↵
7. Watson, Joseph L., David Juergens, Nathaniel R. Bennett, Brian L. Trippe, Jason Yim, Helen E. Eisenach, Woody Ahern, et al. “De Novo Design of Protein Structure and Function with RFDiffusion.” *Nature* 620, no. 7976 (2023):1089–100. <https://doi.org/10.1038/s41586-023-06415-8>. ↵
8. Zheng, Zhiling, Zichao Rong, Nakul Rampal, Christian Borgs, Jennifer T. Chayes, and Omar M. Yaghi. “A GPT-4 Reticular Chemist for Guiding MOF Discovery.” *Angewandte Chemie International Edition* 62, no. 46 (2023):e202311983. <https://doi.org/10.1002/anie.202311983>. ↵
9. Zheng, Zhiling, Oufan Zhang, Christian Borgs, Jennifer T. Chayes, and Omar M. Yaghi. “ChatGPT Chemistry Assistant for Text Mining and Prediction of MOF Synthesis.” *Journal of the American Chemical Society* 145, no. 32 (2023):18048–62. <https://doi.org/10.1021/jacs.3c05819>. ↵
10. Boiko, Daniil A., Robert MacKnight, Ben Kline, and Gabe Gomes. “Autonomous chemical research with large language models.” *Nature* 624, no. 7992 (December 2023):570–8. <https://doi.org/10.1038/s41586->

[023-06792-0](#). ↵

11. Koscher, Brent, Richard B. Canty, Matthew A. McDonald, Kevin P. Greenman, Charles J. McGill, Camille L. Bilodeau, Wengong Jin, et al. “Autonomous, Multi-Property-Driven Molecular Discovery: From Predictions to Measurements and Back.” *Science* 382, no. 6677 (2023): eadi1407.

<https://doi.org/10.1126/science.adi1407>. ↵

12. Strieth-Kalthoff, Felix, Han Hao, Vandana Rathore, Joshua Derasp, Théophile Gaudin, Nicholas H. Angello, Martin Seifrid, et al. “Delocalized, Asynchronous, Closed-Loop Discovery of Organic Laser Emitters.” *ChemRxiv* (2023). <https://doi.org/10.26434/chemrxiv-2023-wqp0d>. ↵

13. Szymanski, Nathan J., Bernardus Rendy, Yuxing Fei, Rishi E. Kumar, Tanjin He, David Milsted, Matthew J. McDermott, et al. “An Autonomous Laboratory for the Accelerated Synthesis of Novel Materials.” *Nature* 624 (November 2023): 86–91. <https://doi.org/10.1038/s41586-023-06734-w>. ↵

14. Sanchez-Lengeling, Benjamin, and Alán Aspuru-Guzik. “Inverse Molecular Design Using Machine Learning: Generative Models for Matter Engineering.” *Science* 361, no. 6400 (2018): 360–5.

<https://doi.org/10.1126/science.aat2663>. ↵

15. Mullin, Rick. “The Tricky Ethics of AI in the Lab.” *Chemical & Engineering News* 101, no. 31 (September 18, 2023). <https://cen.acs.org/business/informatics/tricky-ethics-AI-lab/101/i31>. ↵

16. Urbina, Fabio, Filippa Lentzos, Cédric Invernizzi, and Sean Ekins. “Dual Use of Artificial-Intelligence-Powered Drug Discovery.” *Nature Machine Intelligence* 4, no. 3 (2022): 189–91.

<https://doi.org/10.1038/s42256-022-00465-9>. ↵

17. Ingraham, John B., Max Baranov, Zak Costello, Karl W. Barber, Wujie Wang, Ahmed Ismail, Vincent Frappier, et al. “Illuminating Protein Space with a Programmable Generative Model.” *Nature* 623, no. 7989 (2023): 1070–78. <https://doi.org/10.1038/s41586-023-06728-8>. ↵

18. Madani, Ali, Ben Krause, Eric R. Greene, Subu Subramanian, Benjamin P. Mohr, James M. Holton, Jose Luis Olmos Jr., et al. 2023. “Large Language Models Generate Functional Protein Sequences across Diverse Families.” *Nature Biotechnology* 41 (8): 1099–1106. <https://doi.org/10.1038/s41587-022-01618-2>. ↵

19. Ross, Jerret, Brian Belgodere, Vijil Chenthamarakshan, Inkit Padhi, Youssef Mroueh, and Payel Das. “Large-Scale Chemical Language Representations Capture Molecular Structure and Properties.” *Nature Machine Intelligence* 4 (2022): 1256–64. <https://doi.org/10.1038/s42256-022-00580-7>. ↵

20. Chanussot, Lowik, Abhishek Das, Siddharth Goyal, Thibaut Lavril, Muhammed Shuaibi, Morgane Riviere, Kevin Tran, et al. “Open Catalyst 2020 (OC20) Dataset and Community Challenges.” *ACS Catalysis* 11, no. 10 (2021): 6059–72. <https://doi.org/10.1021/acscatal.0c04525>. ↵

21. Godwin, Jonathan, Michael Schaarschmidt, Alexander Gaunt, Alvaro Sanchez-Gonzalez, Yulia Rubanova, Petar Veličković, James Kirkpatrick, and Peter Battaglia. “Simple GNN Regularisation for 3D Molecular Property Prediction & Beyond.” *arXiv* (June 2021). <https://doi.org/10.48550/arXiv.2106.07971>. ↵
22. Bran, Andres M., Sam Cox, Oliver Schilter, Carlo Baldassari, Andrew D. White, and Philippe Schwaller. “ChemCrow: Augmenting Large-Language Models with Chemistry Tools.” *arXiv* (2023). <https://doi.org/10.48550/arXiv.2304.05376>. ↵
23. Burger, Benjamin, Phillip M. Maffettone, Vladimir V. Gusev, Catherine M. Aitchison, Yang Bai, Xiaoyan Wang, Xiaobo Li, et al. “A mobile robotic chemist.” *Nature* 583, no. 7815 (July 2020):237–41. <https://doi.org/10.1038/s41586-020-2442-2>. ↵
24. Canty, Richard B., Brent A. Koscher, Matthew A. McDonald, and Klavs F. Jensen. “Integrating Autonomy into Automated Research Platforms.” *Digital Discovery* 2, no. 5 (2023):1259–68. <https://doi.org/10.1039/D3DD00135K>. ↵
25. MacLeod, B. P., F. G. L. Parlane, T. D. Morrissey, F. Häse, L. M. Roch, K. E. Dettelbach, R. Moreira, et al. “Self-Driving Laboratory for Accelerated Discovery of Thin-Film Materials.” *Science Advances* 6, no. 20 (2020):eaaz8867. <https://doi.org/10.1126/sciadv.aaz8867>. ↵
26. Steiner, Sebastian, Jakob Wolf, Stefan Glatzel, Anna Andreou, Jarosław M. Granda, Graham Keenan, Trevor Hinkley, et al. “Organic Synthesis in a Modular Robotic System Driven by a Chemical Programming Language.” *Science* 363, no. 6423 (2019):eaav2211. <https://doi.org/10.1126/science.aav2211>. ↵
27. Wu, Tony C, Andrés Aguilar-Granda, Kazuhiro Hotta, Sahar Alasvand Yazdani, Robert Pollice, Jenya Vestfrid, Han Hao, et al. “A Materials Acceleration Platform for Organic Laser Discovery (Adv. Mater. 6/2023).” *Advanced Materials* 35, no. 6 (2023):2370042. <https://doi.org/10.1002/adma.202370042>. ↵
28. Gao, Wenhao, Priyanka Raghavan, and Connor W. Coley. 2022. “Autonomous Platforms for Data-Driven Organic Synthesis.” *Nature Communications* 13, no. 1 (2022):1075. <https://doi.org/10.1038/s41467-022-28736-4>. ↵
29. Saikin, Semion K., Christoph Kreisbeck, Dennis Sheberla, Jill S. Becker, and Alán Aspuru-Guzik. “Closed-Loop Discovery Platform Integration Is Needed for Artificial Intelligence to Make an Impact in Drug Discovery.” *Expert Opinion on Drug Discovery* 14, no. 1 (2019):1–4. <https://doi.org/10.1080/17460441.2019.1546690>. ↵
30. Seifrid, Martin, Robert Pollice, Andrés Aguilar-Granda, Zamyra Morgan Chan, Kazuhiro Hotta, Cher Tian Ser, Jenya Vestfrid, Tony C. Wu, and Alán Aspuru-Guzik. “Autonomous Chemical Experiments: Challenges and Perspectives on Establishing a Self-Driving Lab.” *Accounts of Chemical Research* 55, no. 17 (2022):2454–66. <https://doi.org/10.1021/acs.accounts.2c00220>. ↵

31. Saebi, Mandana, Bozhao Nan, John E. Herr, Jessica Wahlers, Zhichun Guo, Andrzej M. Zurański, Thierry Kogej, et al. “On the Use of Real-World Datasets for Reaction Yield Prediction.” *Chemical Science* 14, no. 19 (2023):4997–5005. <https://doi.org/10.1039/D2SC06041H>. ↵
32. Tshitoyan, Vahe, John Dagdelen, Leigh Weston, Alexander Dunn, Ziqin Rong, Olga Kononova, Kristin A. Persson, Gerbrand Ceder, and Anubhav Jain. “Unsupervised Word Embeddings Capture Latent Knowledge from Materials Science Literature.” *Nature* 571, no. 7763 (2019):95–8. <https://doi.org/10.1038/s41586-019-1335-8>. ↵
33. Ansari, Talha Qasim, Haitao Huang, and San-Qiang Shi. “Phase Field Modeling for the Morphological and Microstructural Evolution of Metallic Materials under Environmental Attack.” *npj Computational Materials* 7, no. 1 (2021):143. <https://doi.org/10.1038/s41524-021-00612-7>. ↵
34. Chen, Chi, Yunxing Zuo, Weike Ye, Xiangguo Li, and Shyue Ping Ong. “Learning properties of ordered and disordered materials from multi-fidelity data.” *Nature Computational Science* 1, no. 1 (January 2021):46–53. <https://doi.org/10.1038/s43588-020-00002-x>. ↵
35. Fare, Clyde, Peter Fenner, Matthew Benatan, Alessandro Varsi, and Edward O. Pyzer-Knapp. “A Multi-Fidelity Machine Learning Approach to High Throughput Materials Screening.” *npj Computational Materials* 8, no. 1 (2022):257. <https://doi.org/10.1038/s41524-022-00947-9>. ↵
36. Greenman, Kevin P., William H. Green, and Rafael Gómez-Bombarelli. “Multi-Fidelity Prediction of Molecular Optical Peaks with Deep Learning.” *Chemical Science* 13, no. 4 (2022):1152–62. <https://doi.org/10.1039/D1SC05677H>. ↵
37. Khoei, A. R., and M. Kianezhad. “A Machine Learning-Based Atomistic-Continuum Multiscale Technique for Modeling the Mechanical Behavior of Ni3Al.” *International Journal of Mechanical Sciences* 239 (February 2023):107858. <https://doi.org/10.1016/j.jmecs.2022.107858>. ↵
38. Montes de Oca Zapiain, David, James A. Stewart, and Rémi Dingreville. “Accelerating Phase-Field-Based Microstructure Evolution Predictions via Surrogate Models Trained by Machine Learning Methods.” *npj Computational Materials* 7, no. 1 (2021):3. <https://doi.org/10.1038/s41524-020-00471-8>. ↵
39. Goldman, Brian, Steven Kearnes, Trevor Kramer, Patrick Riley, and W. Patrick Walters. “Defining Levels of Automated Chemical Design.” *Journal of Medicinal Chemistry* 65, no. 10 (2022):7073–87. <https://doi.org/10.1021/acs.jmedchem.2c00334>. ↵
40. Walters, W. Patrick, and Mark Murcko. “Assessing the Impact of Generative AI on Medicinal Chemistry.” *Nature Biotechnology* 38, no. 2 (2020):143–5. <https://doi.org/10.1038/s41587-020-0418-2>. ↵

41. Zhavoronkov, Alex, and Alán Aspuru-Guzik. “Reply to ‘Assessing the Impact of Generative AI on Medicinal Chemistry.’” *Nature Biotechnology* 38, no. 2 (2020):146. <https://doi.org/10.1038/s41587-020-0417-3>. ↵
42. Ceder, Gerbrand. “Regarding Our Recent A-Lab Article.” LinkedIn, December 1, 2023. <https://www.linkedin.com/pulse/regarding-our-recent-a-lab-article-gerbrand-ceder-0sz6c/>. ↵
43. Palgrave, Robert. 2023. “This Exciting Paper Shows AI Design of Materials, Robotic Synthesis. 10s of New Compounds in 17 Days. But Did They? This Paper Has Very Serious Problems in Materials Characterisation. In My View It Should Never Have Got near Publication. Hold on Tight Let’s Take a Look 🤖.” Twitter, November 30, 2023. https://twitter.com/Robert_Palgrave/status/1730358675523424344. ↵
44. Corrêa, Nicholas Kluge, Camila Galvão, James William Santos, Carolina Del Pino, Edson Pontes Pinto, Camila Barbosa, Diogo Massmann, et al. “Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance.” *Patterns (New York, N.Y.)* 4, no. 10 (October 2023):100857. <https://doi.org/10.1016/j.patter.2023.100857>. ↵
45. Theben, Alexandra, Laura Gunderson, and Laura López-Forés. “Challenges and Limits of an Open Source Approach to Artificial Intelligence.” European Parliament, April 30, 2021. [https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU\(2021\)662908](https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2021)662908). ↵
46. Gómez-Bombarelli, Rafael, Jennifer N. Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, Ryan P. Adams, and Alán Aspuru-Guzik. “Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules.” *ACS Central Science* 4, no. 2 (2018):268–76. <https://doi.org/10.1021/acscentsci.7b00572>. ↵
47. Olivecrona, Marcus, Thomas Blaschke, Ola Engkvist, and Hongming Chen. “Molecular De-Novo Design through Deep Reinforcement Learning.” *Journal of Cheminformatics* 9, no. 1 (2017):48. <https://doi.org/10.1186/s13321-017-0235-x>. ↵
48. Xie, Tian, Xiang Fu, Octavian-Eugen Ganea, Regina Barzilay, and Tommi Jaakkola. “Crystal Diffusion Variational Autoencoder for Periodic Material Generation.” *arXiv* (2022). <https://doi.org/10.48550/arXiv.2110.06197>. ↵
49. Zeni, Claudio, Robert Pinsler, Daniel Zügner, Andrew Fowler, Matthew Horton, Xiang Fu, Sasha Shysheya, et al. “MatterGen: A Generative Model for Inorganic Materials Design.” *arXiv* (2023). <http://arxiv.org/abs/2312.03687>. ↵
50. Subramanian, Akshay, Kevin P. Greenman, Alexis Gervais, Tzuhsiung Yang, and Rafael Gómez-Bombarelli. “Automated Patent Extraction Powers Generative Modeling in Focused Chemical Spaces.”

- Digital Discovery* 2, no. 4 (2023):1006–15. <https://doi.org/10.1039/D3DD00041A>. ↵
51. Gao, Wenhao, Tianfan Fu, Jimeng Sun, and Connor W. Coley. “Sample Efficiency Matters: A Benchmark for Practical Molecular Optimization.” *arXiv* (2022). <https://doi.org/10.48550/arXiv.2206.12411>.
↵
52. Li, Han, Ruotian Zhang, Yaosen Min, Dacheng Ma, Dan Zhao, and Jianyang Zeng. “A Knowledge-Guided Pre-Training Framework for Improving Molecular Representation Learning.” *Nature Communications* 14, no. 1 (2023):7568. <https://doi.org/10.1038/s41467-023-43214-1>. ↵
53. Li, Xinhao, and Denis Fourches. “Inductive Transfer Learning for Molecular Activity Prediction: Next-Gen QSAR Models with MolPMoFiT.” *Journal of Cheminformatics* 12, no. 1 (2020):27. <https://doi.org/10.1186/s13321-020-00430-x>. ↵
54. Zaidi, Sheheryar, Michael Schaarschmidt, James Martens, Hyunjik Kim, Yee Whye the, Alvaro Sanchez-Gonzalez, Peter Battaglia, Razvan Pascanu, and Jonathan Godwin. “Pre-Training via Denoising for Molecular Property Prediction.” *arXiv* (2022). <http://arxiv.org/abs/2206.00133>. ↵
55. Frey, Nathan C., Ryan Soklaski, Simon Axelrod, Siddharth Samsi, Rafael Gómez-Bombarelli, Connor W. Coley, and Vijay Gadepally. “Neural Scaling of Deep Chemical Models.” *Nature Machine Intelligence* 5, no. 11 (2023):1297–305. <https://doi.org/10.1038/s42256-023-00740-3>. ↵
56. Merchant, Amil, Simon Batzner, Samuel S. Schoenholz, Muratahan Aykol, Gwooon Cheon, and Ekin Dogus Cubuk. “Scaling Deep Learning for Materials Discovery.” *Nature* 624 (November 2023):80–5. <https://doi.org/10.1038/s41586-023-06735-9>. ↵
57. Geiger, Mario, and Tess Smidt. “E3nn: Euclidean Neural Networks.” *arXiv* (2022). <https://doi.org/10.48550/arXiv.2207.09453>. ↵
58. Anderson, Brandon, Truong-Son Hy, and Risi Kondor. “Cormorant: Covariant Molecular Neural Networks.” *arXiv* (2019). <https://doi.org/10.48550/arXiv.1906.04015>. ↵
59. Batzner, Simon, Albert Musaelian, Lixin Sun, Mario Geiger, Jonathan P. Mailoa, Mordechai Kornbluth, Nicola Molinari, Tess E. Smidt, and Boris Kozinsky. “E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials.” *Nature Communications* 13, no. 1 (May 2022):2453. <https://doi.org/10.1038/s41467-022-29939-5>. ↵
60. Gasteiger, Johannes, Florian Becker, and Stephan Günnemann. “GemNet: Universal Directional Graph Neural Networks for Molecules.” *arXiv* (2022). <https://doi.org/10.48550/arXiv.2106.08903>. ↵
61. Haghighatlari, Mojtaba, Jie Li, Xingyi Guan, Oufan Zhang, Akshaya Das, Christopher J. Stein, Farnaz Heidar-Zadeh, et al. “NewtonNet: A Newtonian Message Passing Network for Deep Learning of Interatomic

- Potentials and Forces.” *Digital Discovery* 1, no. 3 (2022):333–43. <https://doi.org/10.1039/D2DD00008C>. ↵
62. Qiao, Zhuoran, Anders S. Christensen, Matthew Welborn, Frederick R. Manby, Anima Anandkumar, and Thomas F. Miller. “Informing Geometric Deep Learning with Electronic Interactions to Accelerate Quantum Chemistry.” *Proceedings of the National Academy of Sciences* 119, no. 31 (2022):e2205221119. <https://doi.org/10.1073/pnas.2205221119>. ↵
63. Schütt, Kristof T., Oliver T. Unke, and Michael Gastegger. “Equivariant Message Passing for the Prediction of Tensorial Properties and Molecular Spectra.” *arXiv* (2021). <https://doi.org/10.48550/arXiv.2102.03150>. ↵
64. Unke, Oliver T., Stefan Chmiela, Michael Gastegger, Kristof T. Schütt, Huziel E. Sauceda, and Klaus-Robert Müller. “SpookyNet: Learning Force Fields with Electronic Degrees of Freedom and Nonlocal Effects.” *Nature Communications* 12, no. 1 (2021):7273. <https://doi.org/10.1038/s41467-021-27504-0>. ↵
65. Wang, Yuyang, Ahmed A. Elhag, Navdeep Jaitly, Joshua M. Susskind, and Miguel Angel Bautista. “Generating Molecular Conformer Fields.” *arXiv* (2023). <https://doi.org/10.48550/arXiv.2311.17932>. ↵
66. Gao, Wenhao, and Connor W. Coley. “The Synthesizability of Molecules Proposed by Generative Models.” *Journal of Chemical Information and Modeling* 60, no. 12 (December 2020):5714–23. <https://doi.org/10.1021/acs.jcim.0c00174>. ↵
67. Bradshaw, John, Brooks Paige, Matt J. Kusner, Marwin Segler, and José Miguel Hernández-Lobato. “A Model to Search for Synthesizable Molecules.” In *Advances in Neural Information Processing Systems*, vol. 32. New York: Curran Associates, Inc, 2019. https://proceedings.neurips.cc/paper_files/paper/2019/hash/46d0671dd4117ea366031f87f3aa0093-Abstract.html. ↵
68. Bradshaw, John, Brooks Paige, Matt J. Kusner, Marwin Segler, and José Miguel Hernández-Lobato. “Barking up the Right Tree: An Approach to Search over Molecule Synthesis DAGs.” In *Advances in Neural Information Processing Systems*, vol. 33, 6852–66. New York: Curran Associates, Inc, 2019. <https://proceedings.neurips.cc/paper/2020/hash/4cc05b35c2f937c5bd9e7d41d3686fff-Abstract.html>. ↵
69. Gao, Wenhao, Rocío Mercado, and Connor W. Coley. 2022. “Amortized Tree Generation for Bottom-Up Synthesis Planning and Synthesizable Molecular Design.” *arXiv* (2022). <https://doi.org/10.48550/arXiv.2110.06389>. ↵
70. Hartenfeller, Markus, Heiko Zettl, Miriam Walter, Matthias Rupp, Felix Reisen, Ewgenij Proschak, Sascha Weggen, Holger Stark, and Gisbert Schneider. “DOGS: Reaction-Driven de Novo Design of Bioactive Compounds.” *PLoS Computational Biology* 8, no. 2 (2012):e1002380. <https://doi.org/10.1371/journal.pcbi.1002380>. ↵

71. Vinkers, H. Maarten, Marc R. de Jonge, Frederik F. D. Daeyaert, Jan Heeres, Lucien M. H. Koymans, Joop H. van Lenthe, Paul J. Lewi, Henk Timmerman, Koen Van Aken, and Paul A. J. Janssen. “SYNOPSIS: SYNthesize and OPTimize System in Silico.” *Journal of Medicinal Chemistry* 46, no. 13 (2003):2765–73. <https://doi.org/10.1021/jm030809x>. ↵
72. Fu, Xiang, Tian Xie, Andrew S. Rosen, Tommi Jaakkola, and Jake Smith. “MOFDiff: Coarse-Grained Diffusion for Metal-Organic Framework Design.” *arXiv* (2023). <https://doi.org/10.48550/arXiv.2310.10732>. ↵
73. Praljak, Niksa, Xinran Lian, Rama Ranganathan, and Andrew L. Ferguson. “ProtWave-VAE: Integrating Autoregressive Sampling with Latent-Based Inference for Data-Driven Protein Design.” *ACS Synthetic Biology* 12, no. 12 (2023):3544–61. <https://doi.org/10.1021/acssynbio.3c00261>. ↵
74. Yao, Zhenpeng, Benjamin Sanchez-Lengeling, N. Scott Bobbitt, Benjamin J. Bucior, Sai Govind Hari Kumar, Sean P. Collins, Thomas Burns, et al. “Inverse Design of Nanoporous Crystalline Reticular Materials with Deep Generative Models.” *Nature Machine Intelligence* 3 (2021):76–86. <https://doi.org/10.1038/s42256-020-00271-1>. ↵
75. Krishna, Rohith, Jue Wang, Woody Ahern, Pascal Sturmfels, Preetham Venkatesh, Indrek Kalvet, Gyu Rie Lee, et al. “Generalized Biomolecular Modeling and Design with RoseTTAFold All-Atom.” *bioRxiv* (2023). <https://doi.org/10.1101/2023.10.09.561603>. ↵
76. Batatia, Ilyes, David P. Kovacs, Gregor Simm, Christoph Ortner, and Gabor Csanyi. “MACE: Higher Order Equivariant Message Passing Neural Networks for Fast and Accurate Force Fields.” *arXiv* (December 2022):11423–36. <https://doi.org/10.48550/arXiv.2206.07697>. ↵
77. Chen, Chi, and Shyue Ping Ong. “A universal graph deep learning interatomic potential for the periodic table.” *Nature Computational Science* 2, no. 11 (November 2022):718–28. <https://doi.org/10.1038/s43588-022-00349-3>. ↵
78. Deng, Bowen, Peichen Zhong, KyuJung Jun, Janosh Riebesell, Kevin Han, Christopher J. Bartel, and Gerbrand Ceder. “CHGNet as a Pretrained Universal Neural Network Potential for Charge-Informed Atomistic Modelling.” *Nature Machine Intelligence* 5, no. 9 (2023):1031–41. <https://doi.org/10.1038/s42256-023-00716-3>. ↵
79. Coley, Connor W., Dale A. Thomas, Justin A. M. Lummiss, Jonathan N. Jaworski, Christopher P. Breen, Victor Schultz, Travis Hart, et al. “A Robotic Platform for Flow Synthesis of Organic Compounds Informed by AI Planning.” *Science* 365, no. 6453 (2019):eaax1566. <https://doi.org/10.1126/science.aax1566>. ↵
80. IBM. “IBM RXN for Chemistry.” Accessed February 2, 2024. <https://rxn.res.ibm.com/rxn/robo-rxn/welcome>. ↵

81. Segler, Marwin H. S., Mike Preuss, and Mark P. Waller. “Planning Chemical Syntheses with Deep Neural Networks and Symbolic AI.” *Nature* 555, no. 7698 (2018):604–10. <https://doi.org/10.1038/nature25978>. ↵
82. Nambiar, Anirudh M. K., Christopher P. Breen, Travis Hart, Timothy Kulesza, Timothy F. Jamison, and Klavs F. Jensen. “Bayesian Optimization of Computer-Proposed Multistep Synthetic Routes on an Automated Robotic Flow Platform.” *ACS Central Science* 8, no. 6 (2022):825–36. <https://doi.org/10.1021/acscentsci.2c00207>. ↵
83. Shields, Benjamin J., Jason Stevens, Jun Li, Marvin Parasram, Farhan Damani, Jesus I. Martinez Alvarado, Jacob M. Janey, Ryan P. Adams, and Abigail G. Doyle. “Bayesian Reaction Optimization as a Tool for Chemical Synthesis.” *Nature* 590, no. 7844 (2021):89–96. <https://doi.org/10.1038/s41586-021-03213-y>. ↵
84. Hirschfeld, Lior, Kyle Swanson, Kevin Yang, Regina Barzilay, and Connor W. Coley. “Uncertainty Quantification Using Neural Networks for Molecular Property Prediction.” *Journal of Chemical Information and Modeling* 60, no. 8 (2020):3770–80. <https://doi.org/10.1021/acs.jcim.0c00502>. ↵
85. Abolhasani, Milad, and Eugenia Kumacheva. “The Rise of Self-Driving Labs in Chemical and Materials Sciences.” *Nature Synthesis* 2, no. 6 (2023):483–92. <https://doi.org/10.1038/s44160-022-00231-0>. ↵
86. Unruh, Davis, Venkata Surya Chaitanya Kolluru, Arun Baskaran, Yiming Chen, and Maria K. Y. Chan. “Theory+AI/ML for Microscopy and Spectroscopy: Challenges and Opportunities.” *MRS Bulletin* 47, no. 10 (2022):1024–35. <https://doi.org/10.1557/s43577-022-00446-8>. ↵
87. Ren, Zhichu, Zekun Ren, Zhen Zhang, Tonio Buonassisi, and Ju Li. “Autonomous Experiments Using Active Learning and AI.” *Nature Reviews Materials* 8, no. 9 (2023):563–64. <https://doi.org/10.1038/s41578-023-00588-4>. ↵
88. Huang, Kexin, Tianfan Fu, Wenhao Gao, Yue Zhao, Yusuf Roohani, Jure Leskovec, Connor W. Coley, Cao Xiao, Jimeng Sun, and Marinka Zitnik. “Artificial Intelligence Foundation for Therapeutic Science.” *Nature Chemical Biology* 18, no. 10 (2022):1033–36. <https://doi.org/10.1038/s41589-022-01131-2>. ↵
89. Huang, Ruili. “A Quantitative High-Throughput Screening Data Analysis Pipeline for Activity Profiling.” In *High-Throughput Screening Assays in Toxicology*, edited by Hao Zhu and Menghang Xia, 111–22. New York: Springer, 2016. https://doi.org/10.1007/978-1-4939-6346-1_12. ↵
90. RCSB. “RCSB PDB: Homepage.” Accessed February 2, 2024. <https://www.rcsb.org/>. ↵
91. Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, et al. “Highly Accurate Protein Structure Prediction with AlphaFold.” *Nature* 596, no. 7873 (2021):583–89. <https://doi.org/10.1038/s41586-021-03819-2>. ↵

92. Huang, Kexin, Tianfan Fu, Wenhao Gao, Yue Zhao, Yusuf Roohani, Jure Leskovec, Connor W. Coley, Cao Xiao, Jimeng Sun, and Marinka Zitnik. “Therapeutics Data Commons: Machine Learning Datasets and Tasks for Drug Discovery and Development.” *arXiv* (2021). <http://arxiv.org/abs/2102.09548>. ↵
93. Kearnes, Steven M., Michael R. Maser, Michael Wlekliński, Anton Kast, Abigail G. Doyle, Spencer D. Dreher, Joel M. Hawkins, Klavs F. Jensen, and Connor W. Coley. 2021. “The Open Reaction Database.” *Journal of the American Chemical Society* 143 (45): 18820–26. <https://doi.org/10.1021/jacs.1c09820>. ↵
94. Baibakova, Viktoriia, Mahmoud Elzouka, Sean Lubner, Ravi Prasher, and Anubhav Jain. “Optical emissivity dataset of multi-material heterogeneous designs generated with automated figure extraction.” *Scientific Data* 9, no. 1 (September 2022):589. <https://doi.org/10.1038/s41597-022-01699-3>. ↵
95. Guo, Jiang, A. Santiago Ibanez-Lopez, Hanyu Gao, Victor Quach, Connor W. Coley, Klavs F. Jensen, and Regina Barzilay. “Automated Chemical Reaction Extraction from Scientific Literature.” *Journal of Chemical Information and Modeling* 62, no. 9 (2022):2035–45. <https://doi.org/10.1021/acs.jcim.1c00284>. ↵
96. Kim, Edward, Kevin Huang, Adam Saunders, Andrew McCallum, Gerbrand Ceder, and Elsa Olivetti. “Materials Synthesis Insights from Scientific Literature via Text Extraction and Machine Learning.” *Chemistry of Materials* 29, no. 21 (2017):9436–44. <https://doi.org/10.1021/acs.chemmater.7b03500>. ↵
97. Kononova, Olga, Haoyan Huo, Tanjin He, Ziqin Rong, Tiago Botari, Wenhao Sun, Vahe Tshitoyan, and Gerbrand Ceder. “Text-Mined Dataset of Inorganic Materials Synthesis Recipes.” *Scientific Data* 6, no. 1 (2019):203. <https://doi.org/10.1038/s41597-019-0224-1>. ↵
98. Mukaddem, Karim T., Edward J. Beard, Batuhan Yildirim, and Jacqueline M. Cole. “ImageDataExtractor: A Tool To Extract and Quantify Data from Microscopy Images.” *Journal of Chemical Information and Modeling* 60, no. 5 (2020):2492–509. <https://doi.org/10.1021/acs.jcim.9b00734>. ↵
99. Schwenker, Eric, Weixin Jiang, Trevor Spreadbury, Nicola Ferrier, Oliver Cossairt, and Maria K.Y. Chan. “EXSCLAIM!: Harnessing Materials Science Literature for Self-Labeled Microscopy Datasets.” *Patterns* 4, no. 11 (2023):100843. <https://doi.org/10.1016/j.patter.2023.100843>. ↵
100. Shetty, Pranav, Arunkumar Chitteth Rajan, Chris Kuenneth, Sonakshi Gupta, Lakshmi Prerana Panchumarti, Lauren Holm, Chao Zhang, and Rampi Ramprasad. “A General-Purpose Material Property Data Extraction Pipeline from Large Polymer Corpora Using Natural Language Processing.” *npj Computational Materials* 9, no. 1 (2023):52. <https://doi.org/10.1038/s41524-023-01003-w>. ↵
101. Subramanian, Akshay, Kevin Cruse, Amalie Trewartha, Xingzhi Wang, A. Paul Alivisatos, and Gerbrand Ceder. “Dataset of Gold Nanoparticle Sizes and Morphologies Extracted from Literature-Mined Microscopy Images.” *arXiv* (2022). <http://arxiv.org/abs/2112.01689>. ↵

102. Swain, Matthew C., and Jacqueline M. Cole. “ChemDataExtractor: A Toolkit for Automated Extraction of Chemical Information from the Scientific Literature.” *Journal of Chemical Information and Modeling* 56, no. 10 (2016):1894–904. <https://doi.org/10.1021/acs.jcim.6b00207>. ↵
103. Novak, Matt. “Supermarket AI Gives Horrifying Recipes For Poison Sandwiches And Deadly Chlorine Gas.” *Forbes*, August 12, 2023. <https://www.forbes.com/sites/mattnovak/2023/08/12/supermarket-ai-gives-horrifying-recipes-for-poison-sandwiches-and-deadly-chlorine-gas/>. ↵
104. Biden, Joseph R. “Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence.” The White House. October 30, 2023. Accessed February 2, 2024. <https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/>. ↵