
# A brief overview of deep generative models and how they can be used to discover new electrode materials

Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Electrochemistry

Review Article

# A brief overview of deep generative models and how they can be used to discover new electrode materials

Anders Hellman

As humankind searches for sustainable energy solutions, the demand for electrochemistry has increased. Thus, new and more advanced electrode materials are required. However, finding electrodes that meet the necessary performance is a challenge. Machine learning models can predict key properties such as catalytic activity and stability with surprisingly good accuracy, thus accelerating the process of evaluating materials. However, in most cases, the same models cannot explain how to generate new material compositions. Here, deep generative models can become very valuable. Although issues related to data availability and understanding how these models work still exist, combining deep generative models with computer simulations and laboratory experiments hold great potential for developing the next generation of electrodes. This short review will show recent progress in using deep generative models in related material fields and stress how these models can accelerate the discovery of electrode materials.

**Addresses**
Chalmers University of Technology, Chalmersplatsen 4, Gothenburg, 41296, Sweden

Corresponding author: Hellman, Anders (anders.hellman@chalmers.se)

**Current Opinion in Electrochemistry** 2025, **49**:101629

This review comes from a themed issue on **Fundamental & Theoretical Electrochemistry (2025)**

Edited by **Jan Rossmeisl** and **Henrik Høgh Kristoffersen**

For a complete overview see the Issue and the Editorial

Available online 6 December 2024

**https://doi.org/10.1016/j.coelec.2024.101629**



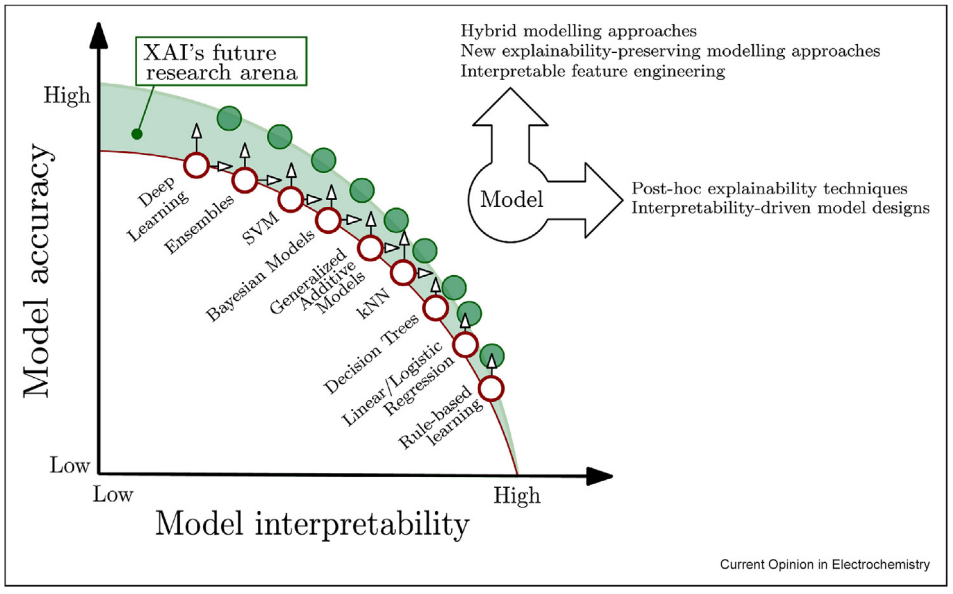
## Introduction

As humankind is becoming more and more concerned about sustainability and energy demands, the interest in electrochemistry for addressing challenges related to energy storage and conversion [1] has become even higher. As the world transitions toward greener energy technologies, there is an urgent need to develop advanced electrochemical materials that must meet strict criteria for performance, durability, and cost-effectiveness to be viable for large-scale applications in the industrial sectors [2].

Despite recent developments in material research, the challenge of finding new or improved materials that meet all these requirements is severe. The traditional paradigm used in material discovery has, to a large extent, relied on experimental trial-and-error, and while often successful, can be very time-consuming and resource-intensive. The introduction of density functional theory (DFT) resulted in a shift of this paradigm, i.e. the possibility of calculating electronic properties of materials from first-principles allowed for the material discovery and design via a computational approach [3–6]. This, combined with advances in high-performance computing, has allowed for high-throughput screening of materials, and as a consequence, there has been a significant speed up in material discovery. However, DFT and other quantum chemistry methods are still limited by the need to know the chemical composition and structural information of materials beforehand, which limits the discovery of novel materials outside of these predefined spaces. To overcome these limitations, methods to predict crystal structures have recently been developed [7,8]. Even with these advancements, the process of finding the "best" material still depends, to a certain degree, on luck, owing to the size of the material space.

Machine learning (ML) has, in recent years, become a popular tool for material science [9–22], as it offers an efficient alternative to the more traditional methods, such as DFT or other quantum chemistry methods. Different ML techniques, see Figure 1, have made it possible to evaluate thousands of material candidates in a fraction of the time required by traditional methods. Here, models such as linear regression, logistic regression, K-nearest neighbors, and Bayesian Models have been widely used in predicting properties of materials and, as a consequence, been able to guide researchers in exploring new material compositions. The combination of high-throughput computational methods with ML have already led to the identification of promising materials for energy storage [23–25] and conversion [26–29], some of which are currently undergoing experimental testing. Despite the progress in using ML in electrochemical material discovery [30–37,2,38,39], some challenges still need to be addressed. For ML

**Figure 1**




The figure illustrates the trade-off between model complexity and interpretability, highlighting where developments in AI may lead to future improvements. This potential for improvement is marked in green, with arrows indicating directions for enhancing both aspects. Reprinted from Publication [12], Copyright (2020), with permission from Elsevier.

models, large, good-quality datasets are needed to make accurate predictions; however, electrochemical systems typically suffer from a lack of consistent and adequate computational/experimental data. To address this, efforts are underway to improve data collection and standardization/sharing across the scientific community [40,39]. A second challenge is that the common use, so far, of ML in electrochemistry is to primarily use ML to evaluate candidates and then identify the best candidates given some criteria, for instance, stability or activity. The initial set of candidates is extracted from various databases, e.g. NOMAD or Open Quantum Materials Database. There are several examples, for instance, Chen et al. [41] and Zhong et al. [42] used ML to identify alloys suitable for HER and electrode materials suitable for $CO_2$ reduction, respectively. Besides identifying the best candidates in a dataset, ML is also used to correlate catalyst descriptors and properties, which later can be used to screen more candidates rapidly. However, examples of generating material candidates in electrochemistry via more advanced ML methods are lacking.

Deep learning models, a subset of ML, are well-suited for material discovery because they can handle complex, high-dimensional data and automatically learn intricate patterns from large datasets without the need for manual feature extraction [43,15]. In materials science, the data often involves many parameters, such as atomic compositions and bonding structures, all resulting in nonlinear relationships that are difficult to model using traditional machine-learning techniques. Deep learning architectures can capture these complexities, allowing for the discovery of hidden correlations in material properties and behaviors [15,17,34,20,44]. However, all of the above ML techniques, including deep learning models, are well-suited to evaluate some properties given a material composition but not the inverse, i.e. suggest a composition with some properties.

Deep generative models build on the power of deep learning and use it to generate novel material compositions and structures [11,25]. One of the key strengths of deep generative models in material discovery is their ability to predict the properties of generated materials. This is made possible as the deep generative models construct a continuous materials vector space, often called latent space. After extracting the essential information from the training dataset and converting it into the latent space, new data points can be generated whenever needed. Moreover, a link between the latent space and a specific property can be established, which allows for the creation of new materials with properties within a desired target range, i.e. enabling inverse design. This predictive power helps narrow the search space significantly, focusing on the most promising candidates. For example, instead of experimentally testing hundreds or thousands of materials, a deep generative model can predict a few top candidates that are likely to meet the performance criteria. Additionally, by examining the latent space constructed by the models, we can gain a better understanding of why certain material compositions outperform others. In the

following sections, we will introduce some of the state-of-the-art deep generative models and highlight some of the latest contributions that have recently been used for material discovery.

## Deep generative models

Deep generative models, such as generative adversarial networks (GANs), variational autoencoders (VAEs), and diffusion models, see Figure 2, have become powerful tools for understanding complex data distributions [15]. They enable the creation of realistic and novel samples and have found applications in various domains, including image generation, natural language processing, and materials science [45–48,24,49]. By learning the underlying distribution of data, these deep generative models can generate new materials that exhibit comparable properties to the original material data [11].

GANs are a type of deep generative model that consists of two neural networks, a generator, and a discriminator, trained in an adversarial manner. The generator takes random noise as input and generates samples that resemble the training data, while the discriminator attempts to distinguish between real data from the training set and the fake data produced by the generator. These two networks are trained simultaneously, with the generator constantly improving its ability to create realistic samples to "fool" the discriminator, and the discriminator learning to improve its ability to distinguish real samples from fake ones. This adversarial dynamic is the foundation of GANs, allowing them to generate highly realistic data. GANs have proven to be highly effective for generating diverse and high-quality materials. In the field of materials science, GANs have been used to create new molecular structures and crystal lattices by learning the distribution of existing compounds [7,2]. For instance, GANs can be trained on a dataset of known organic molecules and then used to generate entirely new molecular structures with desirable properties, such as high thermal conductivity or optimized electronic characteristics. Conditional GANs (cGANs) can be used to generate materials conditioned on specific material properties, making the material discovery process even more targeted [11], which is invaluable for fields like catalyst design and battery technology development.

VAEs are probabilistic generative models that learn input data distribution using latent variables. VAEs comprise two main components: an encoder and a decoder. The encoder reduces the input data to a lower-dimensional form in a latent space while preserving the most important features. Unlike traditional autoencoders, VAEs map the input to a probability distribution, typically a Gaussian distribution, rather than a single point in the latent space. The decoder reconstructs the original data by sampling a point from the latent distribution. This allows for the generation of new samples by sampling different points in the latent space. An important feature of VAEs is that the model learns a continuous latent space, which enables smooth interpolation between different points in the latent space. This implies that new samples can be generated with varying degrees of similarity to the original data. VAEs constructs the latent space by minimizing a loss function that combines two components: a reconstruction loss, which measures how well the reconstructed output matches the original input, and the Kullback–Leibler (KL) divergence, which ensures that the learned latent space conforms to a known distribution, typically Gaussian. After training, the model's latent space can be sampled to generate novel compounds with specific desired properties, such as optimal stability or catalytic activity. This approach has been successfully applied to generate polymer structures, design drug-like molecules, and identify inorganic compounds [11,50,51,24].

Diffusion models are a type of generative model that creates new data by simulating a diffusion process [52,53]. The model gradually changes a simple initial distribution, like Gaussian noise, into a complex data distribution that matches the training data [54]. Diffusion models work in two phases: forward diffusion and reverse diffusion. In the forward diffusion phase, noise is slowly added to the data over a series of steps until the original data are transformed into random noise, diffusing the original information into noise. In the reverse diffusion phase, the model learns a series of transformations to retrieve the original data from the noisy representation, reversing the diffusion process step by step. Diffusion models estimate the gradient of the data distribution with respect to noise, which allows the model to generate new data points by iteratively denoising random noise. This method is very effective at creating high-quality data with complex structures, making diffusion models well-suited for generating novel materials, including the three-dimensional atomic structures. This allows for the discovery of novel compounds with specific physical or chemical properties.

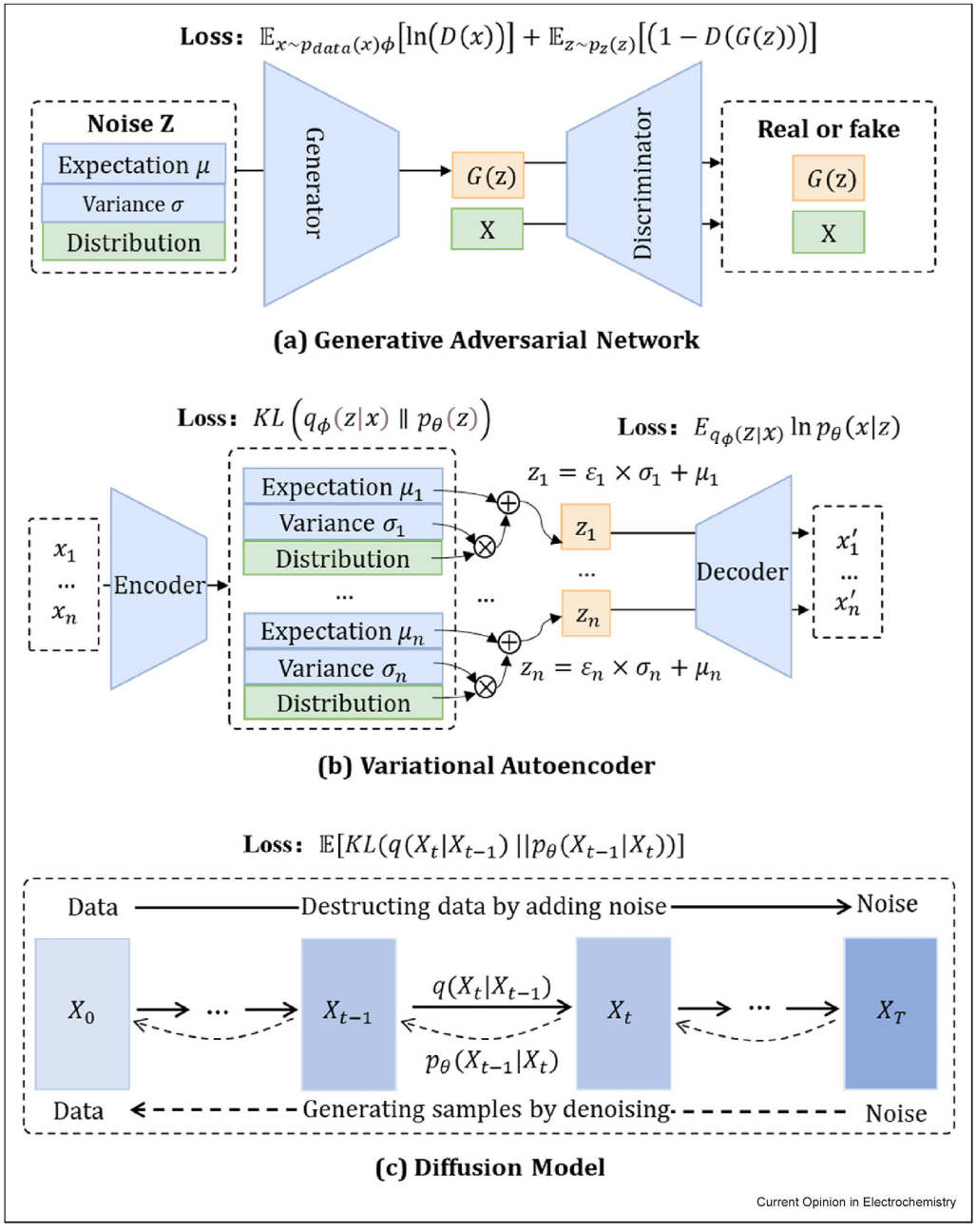
The application of these deep generative models to material discovery, however, also poses specific challenges. Perhaps the most serious challenge is related to data quality and availability. Training deep generative models, as all ML models, requires large amounts of high-quality data, and in materials science, obtaining such data can be challenging. Here, the effort to standardize how data should be reported is important [39]. Ensuring that the generated materials possess these properties requires coupling generative models with property prediction models, which can make the training process more complex. Nonetheless, deep generative models provide numerous opportunities for materials generation. For instance, deep generative models are extremely suited for high-throughput

screening, as the models can propose materials with specific properties. Thus, this approach holds the promise of accelerating the process of materials discovery [55].

## Outlook

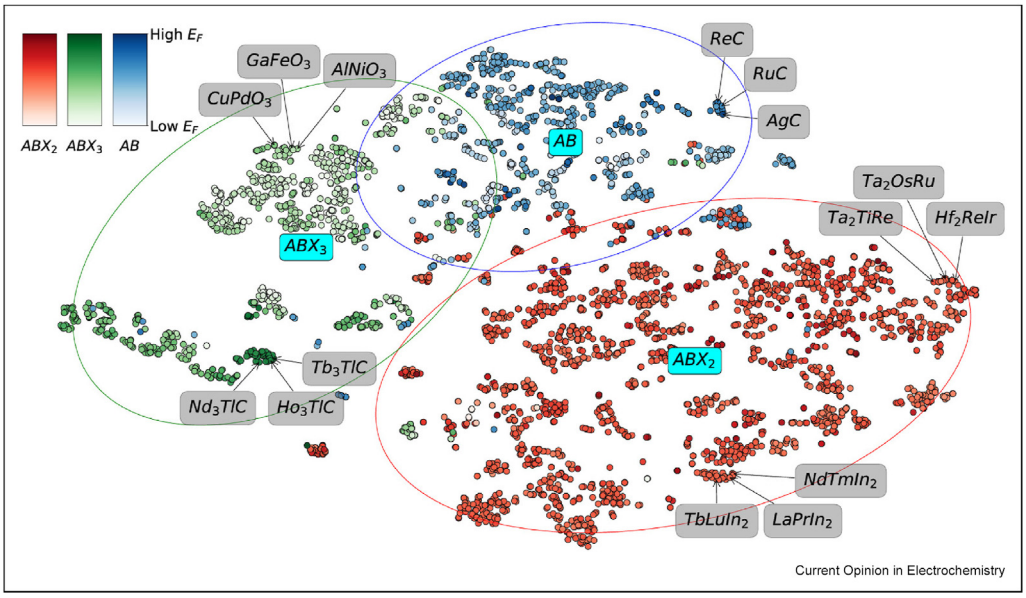
The ability of deep generative models to explore the vast, uncharted design space of materials makes them particularly powerful for discovering new classes of materials [48,46,47]. For example, researchers have used these models to explore the design space of novel molecules [9,47], superconductors [56,23], catalysts for chemical reactions [46,57], and advanced materials for energy storage [23,58]. To illustrate the effectiveness of deep generative models, Figure 3 presents the findings from Court et al. (2020) [51], focusing on cubic binary alloys, ternary perovskites, and Heusler compounds.

**Figure 2**




Schematic picture of the architecture of three types of deep generative models, namely generative adversarial networks (GANs), variational autoencoders (VAEs), and diffusion models. Reprinted from Publication [34], Copyright (2023), with permission from Elsevier.

**Figure 3**



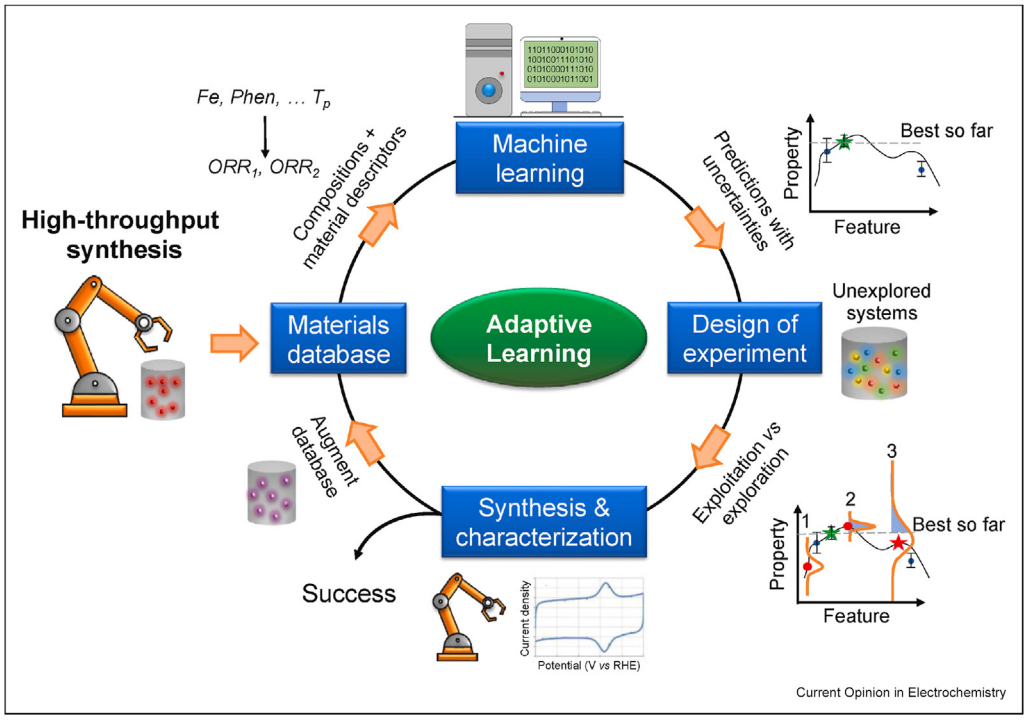

The latent space organizes into clusters based on three distinct structural categories. This arrangement highlights clear groupings driven by both chemical composition and formation energy characteristics. Reprinted from Publication [51] Copyright © 2020 American Chemical Society. This publication is licensed under CC-BY.

The latent space shows clustering according to the three structural types. This clustering of the materials helps to find new materials with similar properties. Furthermore, thanks to the continuous latent space, the deep generative model can suggest novel materials by exploring points between known materials.

The true benefit of deep generative models lies in their ability to suggest material compositions and structures that are not easily imagined by human researchers. In the field of electrode material discovery, this means that deep generative models have the ability to design new materials with optimal properties for specific electrochemical applications, such as battery materials, catalysts for fuel cells, or electrolyzers. Despite these capabilities, the application of deep generative models in electrochemistry has been limited. One reason is the training of ML models, including deep generative models, requires large amounts of high-quality data, something that is often scarce in electrochemical research. As a result, the lack of data hampers the ability to train generative models that could produce realistic and useful predictions in the electrochemical domain. However, the recent development of various databases with standardized entries, e.g. NOMAD or Open Quantum Materials Database, allows for faster access to high-quality data that can be used to train the ML models. Another challenge is the interpretability of deep generative models. Unlike regression or tree-based algorithms, which have clear relationships between input features and output predictions, deep generative models operate through compact latent spaces, which makes it more difficult to understand the reasoning behind a newly generated material composition. Unfortunately, the lack of straightforward interpretation is particularly critical when designing electrode materials, where even minor modifications in material composition can affect their performance.

Another challenge in using deep generative models for computational material discovery is the requirement to create and test the predicted materials in the real world. Even if deep generative models predict materials with good theoretical performance, these materials must be manufactured and tested to confirm their practical functionality. Thus, it is clear that although deep generative models have the ability to create new chemical compositions by exploring complex design spaces, rapid experimental testing is necessary for the output from the deep generative models to be sufficiently exploited. A promising solution to this issue is the use of self-driving laboratories [31,60,59,35,37,44]. These labs employ deep generative models alongside robots to automate the testing and synthesis of materials, see Figure 4. By doing so, they can test materials as quickly as deep generative models generate them, creating a feedback loop that accelerates the discovery of new materials. This approach will help bridge the gap between theoretical discoveries and practical applications.

**Figure 4**




The adaptive learning design loop uses existing data to train machine learning models, predict outcomes, and recommend new samples for testing. The goal is to discover materials with desired properties while minimizing model uncertainty. Reprinted from Publication [59], Copyright (2023), with permission from Elsevier.

## Conclusion

It is clear that the future of material discovery in electrochemistry lies in the continued development and integration of deep generative models with experimental and computational methods. Here, the creation of large, standardized datasets and the development of more interpretable deep generative models will be important. Furthermore, for deep generative models to be effective in material discovery, their predictions must be confirmed via experiment; hence, combining deep generative models with self-driving laboratories—where robots automate synthesis and testing—can establish a feedback loop, which would enable a faster material discovery needed to find the sustainable energy solutions necessary for humankind.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest


1. Sherrell Peter C, Iesalnieks Mairis, Ehrnst Yemima, Rezk Amgad R, Åutka Andris: **Electrocatalysis for green(er) chemistry: limitations and opportunities with traditional and emerging characterization methods for tangible societal impact**. *Advanced Energy and Sustainability Research* 2024, **5**, 2400008.
2. Zhang Chao, Cheng Jun, Chen Yiming, Chan Maria KY, Cai Qiong, Carvalho Rodrigo P, Marchiori Cleber FN, Brandell Daniel, Araujo C Moyses, Chen Ming, Ji Xiangyu, Feng Guang, Goloviznina Kateryna, Serva Alessandra, Salanne Mathieu, Mandai Toshihiko, Hosaka Tomooki, Alhanash Mirna, Johansson Patrik, Qiu Yun-Ze, Xiao Hai, Eikerling Michael, Jinnouchi Ryosuke, Melander Marko M, Kastlunger Georg, Bouzid Assil, Pasquarello Alfredo, Shin Seung-Jae, Kim Minho M, Kim Hyungjun, Schwarz Kathleen, Sundararaman Ravishankar: **Roadmap on molecular modelling of electrochemical energy materials**. *J Phys: Energy* 2023, **5**, 041501. 2023.
3. Nørskov JK, Bligaard T, Rossmeisl J, Christensen CH: **Towards the computational design of solid catalysts**. *Nat Chem* 2009, **1**:37–46.
4. Seh Zhi Wei, Kibsgaard Jakob, Dickens Colin F, Chorkendorff Ib, Nørskov Jens K, Jaramillo Thomas F: **Combining theory and experiment in electrocatalysis: insights into materials design**. *Science* 2017, **355**, eaad4998.
5. Chen Benjamin WJ, Xu Lang, Mavrikakis Manos: **Computational methods in heterogeneous catalysis**. *Chem Rev* 2021, **121**: 1007–1048. PMID: 33350813.
6. Miao Licheng, Jia Wenqi, Cao Xuejie, Jiao Lifang: **Computational chemistry for water-splitting electrocatalysis**. *Chem Soc Rev* 2024, **53**:2771–2807.
7. Kim Sungwon, Noh Juhwan, Gu Geun Ho, Aspuru-Guzik Alan, Jung Yousung: **Generative adversarial networks for crystal structure prediction**. *ACS Cent Sci* 2020, **6**:1412–1420. PMID: 32875082.

8. Zhao Yong, Dilanga Siriwardane Edirisuriya M, Wu Zhenyao, Fu Nihang, Al-Fahdi Mohammed, Hu Ming, Hu Jianjun: **Physics guided deep learning for generative design of crystal materials with symmetry constraints**. *npj Comput Mater* 2023, **9**:38.

9. Butler Keith T, Davies Daniel W, Cartwright Hugh, Isayev Olexandr, Walsh Aron: **Machine learning for molecular and materials science**. *Nature* 2018, **559**:547–555.

10. Kitchin John R: **Machine learning in catalysis**. *Nat Catal* 2018, **1**:230–232.

11. Sanchez-Lengeling Benjamin, Aspuru-Guzik Alan: **Inverse molecular design using machine learning: generative models for matter engineering**. *Science* 2018, **361**:360–365.

12. Barredo Arrieta Alejandro, Diaz-Rodriguez Natalia, Ser Javier Del, Bennetot Adrien, Tabik Siham, Barbado Alberto, Garcia Salvador, Gil-Lopez Sergio, Molina Daniel, Benjamins Richard, Raja Chatila, Herrera Francisco: **Explainable artificial intelligence (xai): concepts, taxonomies, opportunities and challenges toward responsible ai**. *Inf Fusion* 2020, **58**:82–115.

13. Tkatchenko Alexandre: **Machine learning for chemical discovery**. *Nat Commun* 2020, **11**:4125.

14. Wang Anthony Yu-Tung, Murdock Ryan J, Kauwe Steven K, Oliynyk Anton O, Gurlo Aleksander, Brgoch Jakoah, Persson Kristin A, Sparks Taylor D: **Machine learning for materials scientists: an introductory guide toward best practices**. *Chem Mater* 2020, **32**:4954–4965.

15. Pilania Ghanshyam: **Machine learning in materials science: from explainable predictions to autonomous design**. *Comput Mater Sci* 2021, **193**, 110360.

16. Karthikeyan Akshaya, Deva Priyakumar U: **Artificial intelligence: machine learning for chemical sciences**. *J Chem Sci* 2021, **134**:2.

17. * Fuhr Addis S, Sumpter Bobby G: **Deep generative models for materials discovery and machine learning-accelerated innovation**. *Frontiers in Materials* 2022, **9**.
Highlights of different material structure representations and their impact on inverse design strategies using variational autoencoders or generative adversarial networks.

18. Schrier Joshua, Norquist Alexander J, Buonassisi Tonio, Brgoch Jakoah: **In pursuit of the exceptional: research directions for machine learning in chemical and materials science**. *J Am Chem Soc* 2023, **145**:21699–21716. PMID: 37754929.

19. Jain Anubhav: **Machine learning in materials research: developments over the last decade and challenges for the future**. *Curr Opin Solid State Mater Sci* 2024, **33**, 101189.

20. * Suvarna Manu, Pérez-Ramírez Javier: **Embracing data science in catalysis research**. *Nat Catal* 2024, **7**:624–635.
Highlighting the key frontiers and knowledge transfer opportunities across different areas of catalysis.

21. Ding Rui, Chen Junhong, Chen Yuxin, Liu Jianguo, Bando Yoshio, Wang Xuebin: **Unlocking the potential: machine learning applications in electrocatalyst design for electrochemical hydrogen energy transformation**. *Chem Soc Rev* 2024, **53**:11390–11461.

22. * Anthony K: **Cheetham and Ram Seshadri. Artificial intelligence driving materials discovery? perspective on the article: scaling deep learning for materials discovery**. *Chem Mater* 2024, **36**:3490–3495.
A critical examination of the extent to which AI has contributed to advancements in material discovery up to the present time.

23. Thakkar Prit, Khatri Sachi, Dobariya Drashti, Patel Darpan, Dey Bishwajit, Singh Alok Kumar: **Advances in materials and machine learning techniques for energy storage devices: a comprehensive review**. *J Energy Storage* 2024, **81**, 110452.

24. Li Wei, Shen Zhong-Hui, Liu Run-Lin, Chen Xiao-Xiao, Guo Meng-Fan, Guo Jin-Ming, Hua Hao, Shen Yang, Liu Han-Xing, Chen Long-Qing, Nan Ce-Wen: **Generative learning facilitated discovery of high-entropy ceramic dielectrics for capacitive energy storage**. *Nat Commun* 2024, **15**:4940.

25. ** Liu Xinxin, Fan Kexin, Huang Xinmeng, Ge Jiankai, Liu Yujie, Kang Haisu: **Recent advances in artificial intelligence boosting materials design for electrochemical energy storage**. *Chem Eng J* 2024, **490**, 151625.
This study introduces a generative learning approach to quickly discover high-entropy dielectrics. It utilizes encoding-decoding latent space for inverse design and introduces an effective method for designing material systems with desired properties.

26. García-Muelas Rodrigo, López Núria: **Statistical learning goes beyond the d-band model providing the thermochemistry of adsorbates on transition metals**. *Nat Commun* 2019, **10**:4687.

27. Seh Zhi Wei, Jiao Kui, Castelli Ivano E: **Artificial intelligence and machine learning in energy storage and conversion**. *Energy Adv* 2023, **2**:1237–1238.

28. * Xu Yuzhi, Ge Jiankai, Ju Cheng-Wei: **Machine learning in energy chemistry: introduction, challenges and perspectives**. *Energy Adv* 2023, **2**:896–921.
The text below discusses the application of machine learning in energy chemistry, covering areas such as organic photovoltaics, perovskites, catalytic reactions, and batteries. It highlights the recent progress in descriptors and algorithms and their role in designing materials and devices. Additionally, it offers insights into the future direction of research in this field and the potential of machine learning to accelerate the development of sustainable energy sources.

29. Kavalsky Lance, Hegde Vinay I, Meredig Bryce, Viswanathan Venkatasubramanian: **A multiobjective closed-loop approach towards autonomous discovery of electrocatalysts for nitrogen reduction**. *Digital Discovery* 2024, **3**:999–1010.

30. Schlexer Lamoureux Philomena, Winther Kirsten T, Antonio Garrido-Torres Jose, Streibel Verena, Zhao Meng, Bajdich Michal, Abild-Pedersen Frank, Bligaard Thomas: **Machine learning for computational heterogeneous catalysis**. *ChemCatChem* 2019, **11**:3581–3601.

31. Rohr Brian, Stein Helge S, Guevarra Dan, Wang Yu, Haber Joel A, Aykol Muratahan, Suram Santosh K, Gregoire John M: **Benchmarking the acceleration of materials discovery by sequential learning**. *Chem Sci* 2020, **11**:2696–2706.

32. Li Xinyu, Chiong Raymond, Hu Zhongyi, Page Alister J: **Low-cost pt alloys for heterogeneous catalysis predicted by density functional theory and active learning**. *J Phys Chem Lett* 2021, **12**:7305–7311. PMID: 34319099.

33. Jienkulsawad P, Wiranarongkorn K, Chen YS, Arpornwichanop A: **Identifying catalyst layer compositions of proton exchange membrane fuel cells through machine-learning-based approach**. *Int J Hydrogen Energy* 2022, **47**:32303–32314.

34. * Liu Yue, Yang Zhengwei, Yu Zhenyao, Liu Zitu, Liu Dahui, Lin Hailong, Li Mingqing, Ma Shuchang, Avdeev Maxim, Shi Siqi: **Generative artificial intelligence and its applications in materials science: current situation and future perspectives**. *Journal of Materiomics* 2023, **9**:798–816.
Highlights the six main challenges for machine learning in the field, which include the need for high-quality data, domain expertise, generalization, interpretability, usability, security, and cost.

35. ** Szymanski Nathan J, Rendy Bernardus, Fei Yuxing, Kumar Rishi E, He Tanjin, Milsted David, McDermott Matthew J, Gallant Max, Cubuk Ekin Dogus, Merchant Amil, Kim Haegyeom, Jain Anubhav, Bartel Christopher J, Persson Kristin, Zeng Yan, Ceder Gerbrand: **An autonomous laboratory for the accelerated synthesis of novel materials**. *Nature* 2023, **624**:86–91.
Demonstrates the benefits of an autonomous lab combining robotics, databases, machine learning, and automation to optimize the synthesis of inorganic materials, showcasing the synergy of computation, AI, and historical knowledge in experimental research.

36. Steinmann Stephan N, Wang Qing, Seh Zhi Wei: **How machine learning can accelerate electrocatalysis discovery and optimization**. *Mater Horiz* 2023, **10**:393–406.

37. Kodera Masanori, Sayama Kazuhiro: **An automatic robot system for machine learning-assisted high-throughput screening of composite electrocatalysts**. *Digital Discovery* 2023, **2**:1683–1687.

38. Zhu Shan, Jiang Kezhu, Chen Biao, Zheng Shijian: **Data-driven design of electrocatalysts: principle, progress, and perspective**. *J Mater Chem A* 2023, **11**:3849–3870.

39. Moses Abraham Bokinala, Jyothirmai Mullapudi V, Sinha Priyanka, Vines Francesc, Singh Jayant K, Illas Francesc: **Catalysis in the digital age: unlocking the power of data with machine learning**. *WIREs Computational Molecular Science* 2024, **14**, e1730.

40. Brorsson Joakim, Rehnberg Viktor, Arvidsson Adam A, Leion Henrik, Mattisson Tobias, Hellman Anders: **Discovery of oxygen carriers by mining a first-principle database**. *J Phys Chem C* 2023, **127**:9437–9451.

41. Chen Letian, Tian Yun, Hu Xu, Yao Sai, Lu Zhengyu, Chen Suya, Zhang Xu, Zhou Zhen: **A universal machine learning framework for electrocatalyst innovation: a case study of discovering alloys for hydrogen evolution reaction**. *Adv Funct Mater* 2022, **32**, 2208418.

42. Zhong Miao, Tran Kevin, Min Yimeng, Wang Chuanhao, Wang Ziyun, Dinh Cao-Thang, De Luna Phil, Yu Zongqian, Rasouli Armin Sedighian, Brodersen Peter, Sun Song, Voznyy Oleksandr, Tan Chih-Shan, Askerka Mikhail, Che Fanglin, Liu Min, Ali Seifitokaldani, Pang Yuanjie, Lo Shen-Chuan, Ip Alexander, Ulissi Zachary, Sargent Edward H: **Accelerated discovery of co2 electrocatalysts using active machine learning**. *Nature* 2020, **581**:178–183.

43. LeCun Yann, Bengio Yoshua, Hinton Geoffrey: **Deep learning**. *Nature* 2015, **521**:436–444.

44. Toyao Takashi: **Toward accelerated discovery of solid catalysts using extrapolative machine learning approach**. *Chem Lett* 2024, **53**, upae163. 08.

45. Regenwetter L, Nobari AH, Ahmed F: **Deep Generative Models in Engineering Design: A Review**. *ASME. J. Mech. Des.* 2022, **144**, 071704.

46. Hanna Turk, Landini Elisabetta, Kunkel Christian, Margraf Johannes T, Reuter Karsten: **Assessing deep generative models in chemical composition space**. *Chem Mater* 2022, **34**:9455–9467.

47. Bilodeau Camille, Jin Wengong, Jaakkola Tommi, Barzilay Regina, Jensen Klavs F: **Generative models for molecular discovery: recent advances and challenges**. *WIREs Computational Molecular Science* 2022, **12**, e1608.

48. ** Dylan M: **Anstine and Olexandr Isayev. Generative models as an emerging paradigm in the chemical sciences**. *J Am Chem Soc* 2023, **145**:8736–8750.
Evaluates popular generative algorithms like generative adversarial networks, variational autoencoders, flow, and diffusion models, highlighting differences, successful applications, and ongoing challenges in using generative modeling for chemical problems.

49. Hu Jianjun, Li Qin, Fu Nihang: **Generative ai for materials discovery: design without understanding**. *Engineering* 2024, **39**:13–17.

50. Pathak Yashaswi, Singh Juneja Karandeep, Varma Girish, Ehara Masahiro, Deva Priyakumar U: **Deep learning enabled inorganic material generator**. *Phys Chem Chem Phys* 2020, **22**:26935–26943.

51. Court Callum J, Yildirim Batuhan, Jain Apoorv, Cole Jacqueline M: **3-d inorganic crystal structure generation and property prediction via representation learning**. *J Chem Inf Model* 2020, **60**:4518–4535.

52. Yang Sherry, Cho KwangHwan, Merchant Amil, Abbeel Pieter, Schuurmans Dale, Mordatch Igor, Dogus Cubuk Ekin: *Scalable diffusion for materials generation*. 2024.

53. Zeni Claudio, Pinsler Robert, Zügner Daniel, Fowler Andrew, Horton Matthew, Xiang Fu, Shysheya Sasha, Crabbé Jonathan, Sun Lixin, Smith Jake, Nguyen Bichlien, Schulz Hannes, Lewis Sarah, Huang Chin-Wei, Lu Ziheng, Zhou Yichi, Yang Han, Hongxia Hao, Li Jielan, Tomioka Ryota, Xie Tian: *Mattergen: a generative model for inorganic materials design*. 2024.

54. Ho Jonathan, Jain Ajay, Abbeel Pieter: **Denoising diffusion probabilistic models**. *Adv Neural Inf Process Syst* 2020, **33**:6840–6851.

55. Pollice Robert, Gomes Gabriel dos Passos, Aldeghi Matteo, Hickman Riley J, Krenn Mario, Lavigne Cyrille, Lindner-D'Addario Michael, Nigam AkshatKumar, Ser Cher Tian, Yao Zhenpeng, Aspuru-Guzik Alan: **Data-driven strategies for accelerated materials design**. *Accounts Chem Res* 2021, **54**:849–860. PMID: 33528245.

56. Wines Daniel, Xie Tian, Choudhary Kamal: **Inverse design of next-generation superconductors using data-driven deep generative models**. *J Phys Chem Lett* 2023, **14**:6630–6638. PMID: 37462366.

57. Margraf Johannes T, Jung Hyunwook, Scheurer Christoph, Reuter Karsten: **Exploring catalytic reaction networks with machine learning**. *Nat Catal* 2023, **6**:112–121.

58. Xue Haibo, Cheng Guanjian, Yin Wan-Jian: **Computational design of energy-related materials: from first-principles calculations to machine learning**. *WIREs Computational Molecular Science* 2024, **14**, e1732.

59. Kort-Kamp Wilton JM, Ferrandon Magali, Wang Xiaoping, Park Jae Hyung, Malla Rajesh K, Ahmed Towfiq, Holby Edward F, Myers Deborah J, Zelenay Piotr: **Adaptive learning-driven high-throughput synthesis of oxygen reduction reaction feânâc electrocatalysts**. *J Power Sources* 2023, **559**, 232583.

60. Hitt Jeremy L, Yoon Dasol, Shallenberger Jeffrey R, Muller David A, Mallouk Thomas E: **High-throughput fluorescent screening and machine learning for feature selection of electrocatalysts for the alkaline hydrogen oxidation reaction**. *ACS Sustainable Chem Eng* 2022, **10**:16299–16312.