Applied Artificial Intelligence in Materials Science and Material Design

Emigdio Chávez-Angel[1], Martin Børstad Eriksen[2, 3], Alejandro Castro-Alvarez[4], Jose H. Garcia[1], Marc Botifoll Moral[1], Oscar Avalos-Ovando[5], Jordi Arbiol[1, 6] and Aitor Mugarza[1, 6]

[1] Catalan Institute of Nanoscience and Nanotechnology, CSIC and BIST, Campus UAB, Bellaterra, 08193 Barcelona, Spain

[2] Institut de Física d'Altes Energies (IFAE), The Barcelona Institute of Science and Technology, 08193 Bellaterra (Barcelona), Spain

[3] Port d'Informació Científica, Campus UAB, C. Albareda s/n, 08193 Bellaterra (Barcelona), Spain

[4] Departamento de Ciencias Preclínicas, Facultad de Medicina, Universidad de La Frontera, Temuco 4811230, Chile

[5] Department of Physics and Astronomy and Nanoscale and Quantum Phenomena Institute, Ohio University, Athens, Ohio 45701, United States.

[6] ICREA Pg. Lluís Companys 23, Barcelona 08010, Catalonia, Spain

* Corresponding author: emigdio.chavez@icn2.cat

Abstract

Materials science has traditionally relied on a combination of experimental techniques and theoretical modelling to discover and develop new materials with desired properties. However, these processes can be time-consuming, resource-intensive, and often limited by the complexity of material systems. The advent of artificial intelligence (AI), particularly machine learning, has revolutionized materials science by offering powerful tools to accelerate the discovery, design, and characterization of novel materials. AI not only enhances the predictive modelling of material properties but also streamlines data analysis in techniques like X-ray diffraction, Raman spectroscopy, scanning probe microscopy and electron microscopy. By leveraging large datasets, AI algorithms can identify patterns, reduce noise, and predict material behaviour with unprecedented accuracy. In this review, we highlight recent advancements in AI applications across various domains of materials science, including spectroscopy, synchrotron studies, scanning probe and electron microscopies, metamaterials, atomistic modelling, molecular design and drug discovery. We discuss how AI-driven methods are reshaping the field, making material discovery more efficient and paving the way for breakthroughs in material design and real-time experimental analysis.

Introduction

Over the past decade, artificial intelligence (AI) has made remarkable advancements, significantly impacting various sectors of society. The term AI can be traced back to the Dartmouth Summer Research Project on Artificial Intelligence in 1956 and the first neural network was trained by Frank Rosenblatt in 1958.[1] Over the years, the field has been through various epochs of optimism followed by pessimism and less interest in the field, when resources vanished and progress grinded to a halt.

A pivotal moment in the field of AI occurred with the application of neural networks (NNs) in the ImageNet challenge.[2] ImageNet is a large and tagged image dataset compiled by Fei-Fei Li's research group.[3] The large size and variety of this dataset presented a sufficiently challenging task for AI models. In the 2012 competition, Geoffrey Hinton's team achieved a significant breakthrough.[4] They achieved higher precision than previously observed, without the need for manual feature engineering. This success was supported by several key advancements. The architecture employed was a convolutional neural network (CNN), which is particularly well-suited for image processing. Additionally, the model was trained using GPUs, which greatly enhanced computational efficiency.

The remarkable improvement in results, along with other technological advances, revived interest in NNs.

The basic form and core functionality of NNs is making predictions. Like in classical machine learning (ML) methods, they can be trained on data to classify or predict quantities. However, NNs are undergoing a fascinating revolution. Making predictions is still the foundation, but this has led to unexpected advances. The generative neural networks (GANs) created by Ian Goodfellow[5] started generative AI and new paradigm "*learn how to teach a network to generate images*".[5] They used two networks: a generator that created images and a discriminator distinguishing between real and generated images. Trained together with the two networks competing, the generator learned to create images. The first examples were blurry, ghost-like faces, but was only the starting point. Generative AI has progressed rapidly, evolving from producing small images to generate highly realistic ones. This progress involved numerous iterations and innovations. After only four years, the BigGAN network could generate images that humans struggled to identify as machine generated.[6] Lately diffusion models, also based on NNs, can generate increasingly complex scenes and produce artistic images.

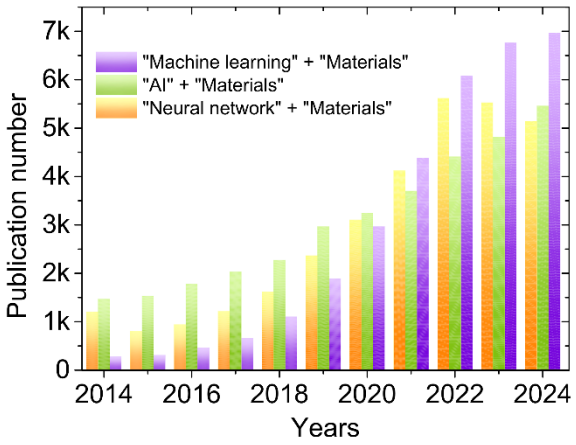
The advancement in AI is also a story about increasing amount of computer power. Early AI efforts were limited by the "slow" computers of the time, and while some researchers hypothesised that greater computing power and larger datasets would drive progress, this remained speculative. Since the advent of AI, GPUs (originally developed for gaming) have become an essential tool to speed up the training of NNs. Their ability to perform parallel operations make them more efficient than CPUs. Moreover, the introduction of the CUDA programming language allowed developers to code directly on GPUs, which played a pivotal role in the training of neural networks with many free parameters. Over time, GPU processing power has significantly advanced, with optimizations such as specialised numerical types. The demand for powerful GPUs has surged as increasingly larger models are trained, leading to exponential growth in the processing power required and even rendering GPUs a scarce resource in recent years.[7,8]

As an example, language models (like ChatGPT) have recently gained popularity, leading to the development of numerous similar models. Although natural language processing (NLP) is a longstanding field, NNs now dominate, with recent advancements building on transformer networks using attention mechanisms. These innovations have garnered significant attention, with large language models (LLM) becoming closely associated with AI. The results are impressive, with LLMs writing text, producing summaries and answering questions with an increasingly sophisticated manner. The state-of-the-art models are trained in large data centres, using 10s of thousands of high-end GPUs, with costs in the tens of millions of dollars.

While large players dominate some subfields needing enormous amounts of computing power, it is possible to perform AI research with more modest resources. An important contribution to lowering the barrier to entry was that Google open-sourced TensorFlow in 2015,[9] followed by the release of PyTorch from Meta in 2016.[10] Although there existed frameworks for training neural networks earlier, these mature frameworks makes applying neural networks and building new methods much simpler. Currently PyTorch is dominating research, but Tensorflow has advantages in large deployment and is also popular for its simple Keras interfaces, which appeals to beginners. The open availability of these tools has facilitated broader participation in AI research, with major tech companies playing a key role in advancing the field by open-sourcing their technologies.

The rapid advances in AI, both in theoretical research and its application to consumer products, have garnered significant interest from scientists across diverse disciplines. Researchers in fields such as biology, optics, and astronomy have increasingly adopted AI methods, leading to a sharp rise in publications with "machine learning", "AI" or "Neural network" in their keywords (see figure 1). Some outstanding advances in a "scientific context" include the application of AI in protein folding through models like AlphaFold[11], which has garnered widespread recognition, including a Nobel Prize in Chemistry. The use of AI in astronomical research for the automated detection and classification of

celestial objects[12,13], and the integration of AI-driven imaging techniques in medical diagnostics, enabling more accurate and efficient disease detection.[14,15] Numerous conferences and workshops now focus specifically on the integration of AI into various fields, as scientists explore how best to use these powerful tools for advancing their research.



Topic / Years	"Machine learning" + "Materials"	"AI" + "Materials"	"Neural network" + "Materials"
2024	6968	5460	5145
2023	6764	4816	5530
2022	6079	4413	5620
2021	4384	3705	4123
2020	2967	3244	3108
2019	1892	2967	2368
2018	1105	2276	1618
2017	661	2037	1217
2016	465	1780	944
2015	311	1535	805
2014	287	1472	1204

Figure 1. Left Number of publications in the last 10 years including the keywords “Machine learning” + “Materials” (violet), “AI” + “Materials” (violet) and “Neural network” + “Materials”. Right Table of number of publications in the with the number of publications in 2024–2014 period. All the data was extracted from web of science.[16]

In material science, the development of new materials relied on a combination of experimental methods and theoretical models, often requiring iterative trial-and-error approaches to optimize material properties. With the increasing complexity of material systems, this process has become increasingly inefficient, prompting a shift toward AI-driven methods that can significantly accelerate discovery and innovation. AI has facilitated the development of predictive models that forecast material behaviour under different conditions, aiding in experimental design and result interpretation.[17] This predictive power is expected to improve the structural-functional correlations, leading to a more cost-effective research cycle and automating materials discovery and synthesis.[18,19] AI tools and DL models are crucial for predicting molecular structures, optimizing fabrication processes, and simulating material behaviours. AI has been applied to identify advanced materials for energy storage[20–22], superconductors[23] and thermoelectric[24,25]. Outstanding research includes AI-based predictions of material properties from chemical composition[26] and the use of unsupervised word embeddings to extract insights from millions scientific abstracts in materials science, physics, and chemistry.[27]

In this review, we explore the various ways AI is being applied to advance materials science. We cover the use of AI in spectroscopic analysis, synchrotron facilities, electron microscopy and scanning probe techniques, where it enhances real-time data processing and analysis, allows autonomous acquisition, tip preparation, among others. Additionally, we examine the role of AI in metamaterial design, atomistic modelling, and drug discovery, highlighting how AI-driven methods are solving long-standing challenges in each area. Finally, we discuss the future directions of AI in materials science, with a particular emphasis on research initiatives like In-situ Correlative Facility for Advanced Energy Materials (InCAEM) project.[28] That is a collaborative project within the Spanish Advanced Materials Programme, which is developing unique infrastructure to enable the correlative use of (S)TEM, SPM (AFM/STM), and synchrotron radiation techniques to address key scientific challenges aligned with the European Green Deal’s goal of fostering a sustainable economy. In addition to instrumentation, InCAEM will develop a robust computational infrastructure for automated AI-based data analysis, essential for managing the large datasets generated by in-situ experiments. This integration will

streamline experimental workflows, enhance data interpretation, and accelerate scientific discovery, advancing our understanding of materials for energy and environmental sustainability.

Applied-AI in Raman and Infrared spectroscopy

Spectroscopic methods, like Infrared or Raman spectroscopy, are fundamental tools in materials science for analysing the structure and composition of materials. These techniques can generate large amounts of data, which often require extensive analysis to extract meaningful information. Traditional methods of data analysis can be time-consuming and may struggle to identify subtle patterns or anomalies within large datasets. For the case of Raman spectroscopy, the fitting of the modes can bring very useful information about the system. The peak position of the Raman band can be used as indicator of local strain[29], temperature[30] or doping concentration[31], the intensity ratio of D and G bands (I_D/I_G) in carbon allotropes is commonly used as an indicator of the purity[32] and its lineshape can be used to determine average sheet size[33], peak position and linewidth can also provide information about phonon anharmonicity and electron-phonon interactions[34], among others. While the fit of individual peaks is straight forward, this task can be more challenging for the case of the analysis of a full Raman map (or set of mapping), a time series measurement or in-situ experiments. Moreover, if during these measurements the background is changing with time, a pretreatment of the whole data set must be carried out before the traditional fitting.

Initially, chemometric methods were the key tools for data analysis and interpretation in spectroscopy.[35] However, with the arrival of AI and ML, the capabilities of data analysis have been dramatically enhanced. This evolution has not only improved the accuracy of spectroscopic investigation but has also opened new possibilities for real-time, in-situ, and predictive analytics. Chemometrics emerged in the 60's as a multidisciplinary field combining chemistry, mathematics, and statistics to extract meaningful information from spectroscopic data.[36] The main aim of chemometrics was to simplify complex spectral data into more interpretable forms, allowing for the qualitative and quantitative analysis of chemical compounds. Key chemometric tools (which are also forms of ML), include principal component analysis (PCA), partial least squares regression (PLS), multivariable curve resolution (MCR), linear discriminant analysis (LDA), k-nearest neighbour (KNN) and k-means. PCA is widely used for reducing the dimensionality of complex spectroscopic data by transforming variables into orthogonal components that capture the maximum variance.[37–39] This unsupervised method is commonly applied in fields such as pharmaceutical analysis[40,41], food authentication[42–44], medical diagnostics[45], and material science[46] due to its simplicity and efficiency in handling large datasets.

PLS addresses cases with more variables than observations, frequently in IR and Raman spectroscopy, by finding latent variables that explain both predictor (X) and response (Y) data, making it ideal for quantitative analysis of complex mixtures.[47] MCR is used to resolve overlapping spectral signals into pure components.[48] LDA is also applied in dimensionality reduction and classification, separating classes by maximizing inter-group variance and minimizing intra-group variance. KNN is used for classification and regression problems. KNN calculates the distance between a data point and known samples to find the K closest sample categories. On the other hand, K-means groups observations into k clusters with the nearest mean or centroid. K-means selects k random centroids, assigns data points to the nearest, updates centroids by averaging cluster points, and repeats until assignments stabilize or the iteration limit is reached.[48]

Figure 2 shows a typical chemometric analysis of a photoluminescence (PL) map using k-means (a and b) and PCA (c and d). The k-means analysis shows automatic clustering of the PL signal, aiding researchers in faster signal analysis and enabling automated examination of individual spectra.

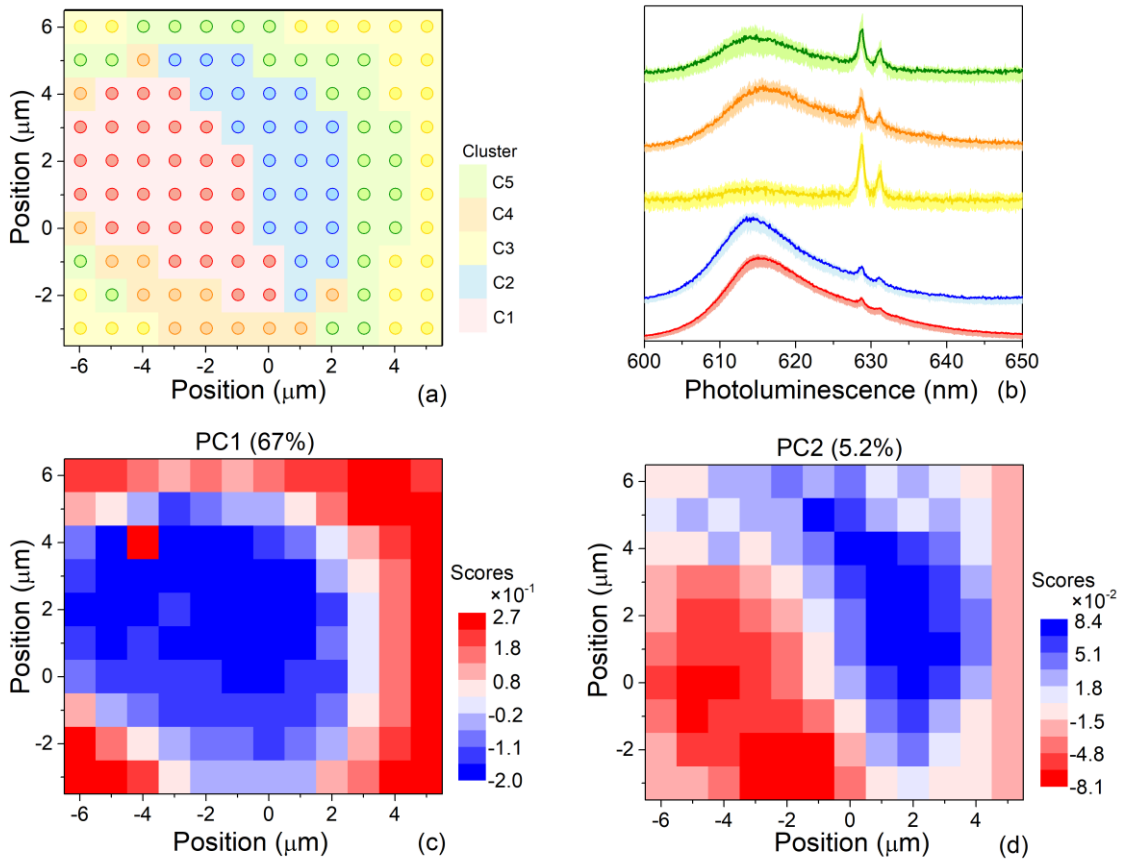


Figure 2. Chemometric analysis of strain distribution of photoluminescence (PL) signal in suspended WS_2 layers. (a) k-mean clusters of PL signal, (b) averaged spectra and standard deviation of individual clusters (shadow areas). First (c) and second (d) PCA of the PL maps. Adapted from[49]

While chemometrics provided foundational tools for the analysis of spectroscopic data, its capabilities are limited to manage large and complex datasets generated by modern instruments and the combinations of spectroscopic tools. These techniques often required substantial domain expertise and manual intervention for calibration and interpretation. In this context, the AI appeared as powerful tool to help with the analysis of spectroscopic data. In spectroscopic analysis ML and DL algorithms has been mainly applied to denoise signal, classification tasks and prediction.

NNs, such as CNNs and encoder-decoder architectures, have demonstrated great potential in removing noise from raw Raman and IR signals.[50–54] This is very useful in biological and chemical analysis, where the raw signal often contains a large amount of noise from environmental sources or background from photoluminescence, fluorescence or plasmonic effects. By using these tools is possible to reconstruct and denoise the raw spectra without the need for complex preprocessing techniques. **Figure 3** presents an example of this application for spectrum denoising. The input consists of numerical spectra with 1 to 7 Gaussian peaks, varying in linewidth and intensity. An arbitrary background (ranging from constant to Gaussian-like) along with white noise, was added to create the noisy input signal. The NNs were trained on this noisy data to produce a clean output. The bottom panel compares the performance of two denoising architectures, demonstrating the effectiveness of these small NNs in removing noise from the data.

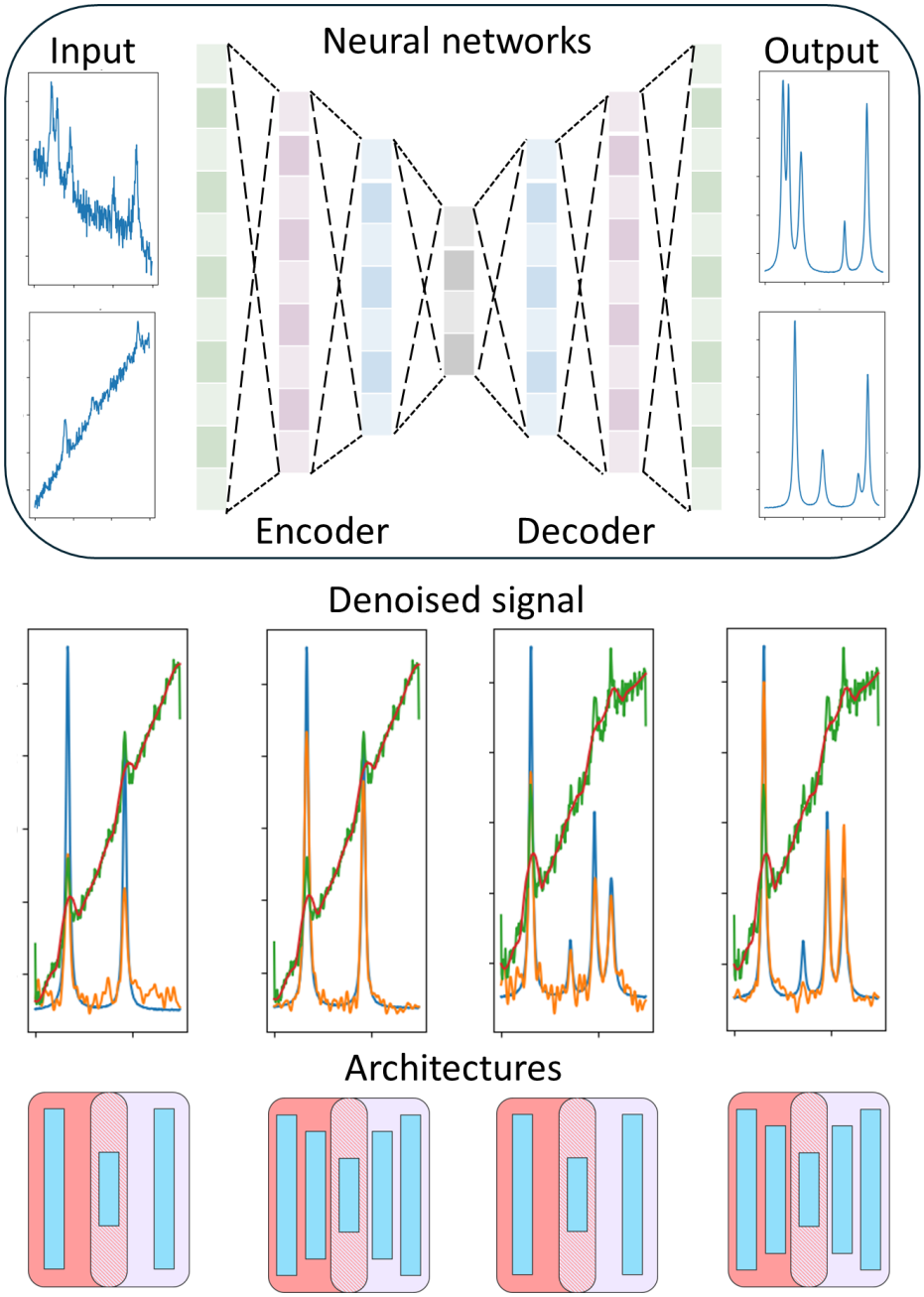


Figure 3. Example of some encoder-decoder architectures for denoising of spectroscopic signal. The NNs are feed using the noise data and the output is a clean signal. Bottom: denoising performance of different architectures, green line inputs, blue lines expected output, orange reconstructed signal.

CNNs have also streamlined spectroscopic data analysis, automatic tasks like peak detection, classification, and quantification. For example, Lim et al.[55] demonstrated the fast detection and classification of the microplastic (MPs) using Raman spectroscopy and CNNs. Similarly, SpecATNet, a self-attention-based NN architecture, was used to process thousands of spectra generated by surface enhanced Raman spectroscopy (SERS).[56] This NN was able to identify and classify different MPs based on the Raman signal. Through the incorporation of a self-attention mechanism, the model focused on relevant aspects of the spectral data, improving its ability to differentiate between various types of MPs, even in the presence of noise or overlapping signals. This automation reduced substantially the

time and effort required for manual analysis, making detection more efficient, faster, and scalable. Recently Yan et al. developed a fingerprint region-based data augmentation (FRDA) model to improve the classification of MPs using FTIR data.[57] FRDA integrates explainable AI (XAI) with CNNs to identify critical spectral regions (fingerprint regions) relevant for classification. These regions are clustered using a Gaussian Mixture Model (GMM) and recombined to generate augmented samples, addressing data imbalance. The model enhances classification accuracy, surpassing traditional augmentation methods (Support Vector Machines, SVM), while providing insights into key spectral regions for MP identification. XAI refers to methods and techniques in AI that make the decision-making processes of ML models more transparent, interpretable, and understandable to the users.[58,59] Unlike traditional "black-box" AI models, which often provide predictions or outputs without explaining the reasoning behind them, XAI aims to clarify how and why a model arrives at its conclusions. Some XAI tools includes shapley additive explanations (SHAP), local interpretable model-agnostic explanations (LIME), class activation mapping (CAM), partial dependence plots (PDP), individual conditional expectation (ICE) among others. They are used to make ML models transparent and interpretable. In spectroscopy, XAI identifies key spectral features, such as bands, peaks or linewidths, that influence AI-driven predictions, improving the reliability and acceptance of these AI models.[58]

In another application, Yoo et al[60] also applied XAI to study phonon-assisted emission in tungsten diselenide (WSe_2) monolayers, using CNNs trained on Raman scattering (RS) and photoluminescence (PL) data. XAI techniques such as PDP and ICE revealed correlations between specific Raman modes and exciton dynamics, identifying the A_1 mode as a critical factor. The same group used CNNs and extreme gradient boosting (XGB) to predict the carbon-to-oxygen (C/O) ratio in graphene oxide (GO). The XAI techniques (PDP and feature importance analysis) identifying the D* Raman mode as a key indicator of GO reduction.[61] The results were also validated through quantum mechanical simulations.

Other studies have applied ML to spectroscopic data for structural analysis. Solis-Fernandez and Ago[62] used several ML models classify and predict the twist angle of a bilayer graphene (BG) based on the intensity and linewidth 2D, G, R and R' Raman modes. They achieved a predictive accuracy ~ 99% using Random Forest (RF) approach. In the same line, Vicent et al.[63] used GMM and PCA to analyse Raman spectra and identify regions with specific twist angles of BG (clustering and classification). GMM was used for data clustering, while PCA helped reduce the dimensionality of the data to make clustering more efficient. Recently, Hu et al. used a combination of GAN and CNNs to classify and predict twist angle of BG achieving an accuracy and recall of 99.9%.[64]

In another novel study, E. Chavez-Angel et al.[65] applied ML (RF and KNN) to predict temperature distributions on poly(methyl methacrylate) (PMMA) thin films using the temperature-dependence of FTIR spectra. The experiment consisted of two steps: first, calibrating the FTIR signal of the polymer as a function of temperature using a cryostat, and second, generating a temperature distribution with a micrometre-scale metal heater. Once calibrated, the FTIR signal was mapped across the surface to monitor local temperature distributions, translating the FTIR map into precise temperature measurements. By employing ML, the study exploited not only individual FTIR peaks but also the entire spectral range, enabling the FTIR spectra to function as a highly effective local thermometer. This work highlights the synergy between ML and FTIR spectroscopy for novel applications.

Open-source tools further support AI-driven spectroscopy. Recently Georgiev et al.[66] introduced RamanSPy, an open-source Python package designed for data analysis of Raman spectra. It offers a suite of modular tools for data loading, preprocessing, analysis, and visualization. RamanSPy aims to facilitate reproducible research by enabling pipeline automation, integration of custom methods, and access to curated datasets for diverse applications such as medical diagnostics, materials science, and environmental studies. Its flexible and user-friendly architecture positions it as a robust platform for advancing AI-driven Raman spectroscopy. Likewise, Orange Data Mining[67–69] developed by Bioinformatics Lab at University of Ljubljana, Slovenia, provides an intuitive drag-and-drop platform

for data analysis tasks such as classification, clustering, and bioinformatics. These tools enhance accessibility and facilitate advanced ML-driven analysis in spectroscopy.

Applied-AI in synchrotron facilities

Synchrotron facilities are at the forefront of scientific research, providing powerful beams lines from UV to IR and X-rays that enable detailed exploration of the atomic and molecular structures of materials. These facilities are used across various fields, including biology, materials science, chemistry, and physics. However, with the increasing volume and complexity of the data produced by synchrotron experiments, researchers are turning to the application of AI to streamline data acquisition, processing, and analysis, accelerating scientific discoveries.[70]

Synchrotrons offer unique opportunities for research but face significant challenges. In advanced techniques like tomography, spectroscopy, and diffraction, managing the volume and complexity of the data generated during experiments can be difficult and time consuming. The analysis and processing also require computational resources, and in some cases, real-time processing. Given the limited availability and high demand for these facilities, beamtime optimization is crucial. This emphasizes the need to maximize experimental efficiency. The management, storage, and sharing of the generated data present logistical and technical challenges that must be addressed.[71,72] In this context, AI-based tools have significantly contributed to the advancement of synchrotron technologies in two main areas: (*i*) *data analysis* and (*ii*) *real-time applications* (See Figure 4).

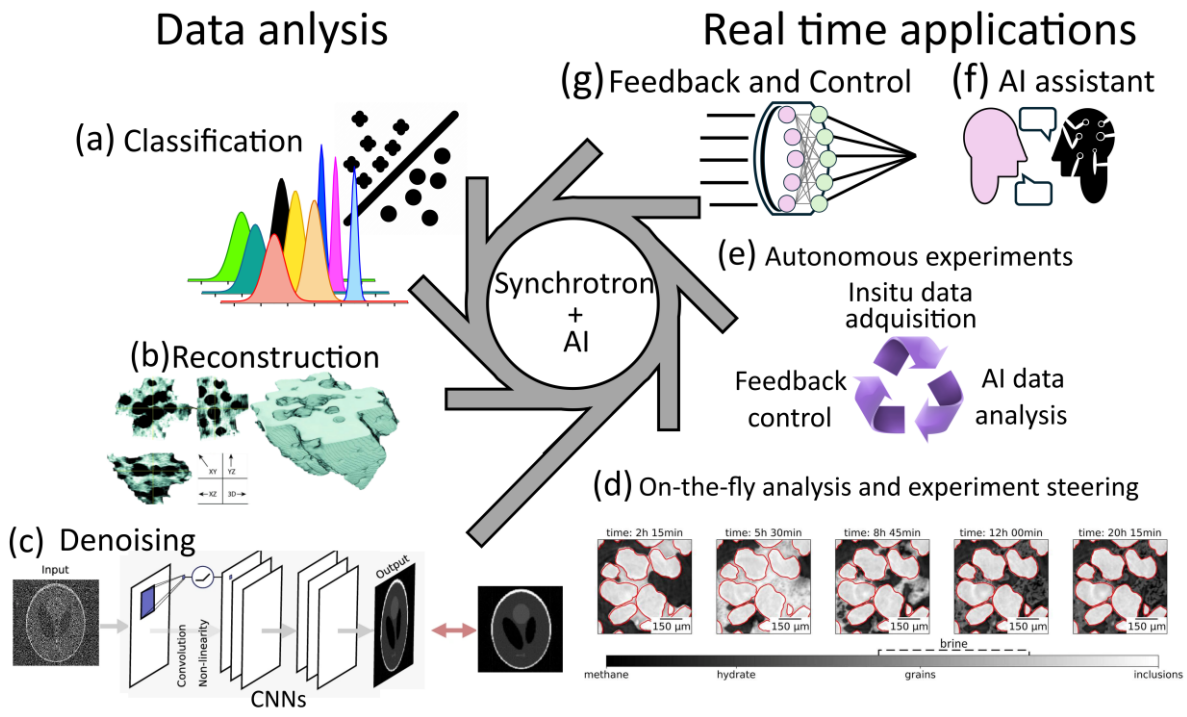


Figure 4. Summary of the main application of AI in synchrotron facilities. Data analysis: (a) automatic classification, (b) tomography reconstruction (adapted with permission from Ref [73]) and (c) denoising (adapted with permission from Ref [74]). Real time applications: (d) on-the-fly analysis (adapted with permission from Ref [75]), (e) Autonomous experiments (adapted from Ref [76]) (f) chatbot assistant and (g) autonomous alignment and management.

Applied-AI in data analysis in synchrotron experiments

The capability of AI to identify complex patterns within massive datasets capability is particularly beneficial in fields such as X-ray spectromicroscopy, in which experiments often generate millions of spectra.[71] One area where AI has made remarkable advances is the classification of crystalline

systems and space groups from X-ray diffraction patterns. Traditionally, this process relies heavily on manual trial-and-error operations, which are time-consuming and error-prone. However, AI algorithms can learn from large data sets of simulated or experimental diffraction patterns and accurately predict the symmetry and space group of a crystal.[77] Suzuki et al. [78] used five ML algorithms to classify crystal system and space group of powder XRD data. The extremely randomized trees (ERT) model achieved a high classification accuracy 92.26% for seven crystal systems and 230 space groups. The analysis of the ERT model revealed interesting insights into the factors influencing crystalline system classification. For example, the model highlighted the importance of the number of diffraction peaks and the positions of the lowest-angle peaks as key indicators of crystal symmetry. This finding is consistent with the intuition of experienced crystallographers, who often rely on these features to make initial estimates of crystal symmetry.

The reconstruction of synchrotron X-ray tomography is another popular application of AI. Despite that the mathematical foundation of image reconstruction is well stablished, the development of reconstruction algorithms is still a challenge due the inherent noise and imperfections in the measurement data.[79] Image-to-image models based on CNNs and GANs has been successfully applied in tomography reconstruction. [79,80] Yang et al.[73] introduced a tomographic reconstruction method using GAN to address challenges like missing-wedge data, which cause distortions in traditional reconstructions. They created GANrec algorithm which generates candidate reconstructions from sinograms and refines them using a physics-based model, eliminating the need for large training datasets. The method showed superior accuracy, particularly with undersampled data, and was validated with real-world tests on a zeolite particle, delivering high-quality reconstructions suitable for quantitative analysis (see **Figure 4b**).

X-ray tomography often produces noisy images due to low dose or time constraints, particularly in dynamic processes. Hendriksen et al. [74,81] developed Noise2Inverse algorithm to obtain high-quality images based on noise images (see **Figure 4c**). This method uses noisy data to train a CNN for denoising across three domains: space (3D micro-tomography), time (dynamic tomography), and spectrum (X-ray diffraction computed tomography). Other example for denoising is TomoGAN developed by Liu et al.[82] They built a denoising algorithm based on GAN to improve the quality of reconstructed X-ray tomography images taken at low-dose conditions. It is important to mention the creation of TomoBank tomographic dataset designed by De Carlos et al. [83] This dataset is a reference library of X-ray tomographic images. Other key problem in tomography is the aligning of the projection images a crucial step before reconstruction. In this line, Yang et al.[84] introduced CNNs to calibrate the centre-of-rotation (CoR) for X-ray tomography. They developed Xlearn a python toolbox library. Nowadays, this library goes beyond its original aim (image classification and image transformation) allowing: correction of instrument and beam instability artifacts, low-dose image enhancement, feature extraction and segmentation.[85]. Other example is the work of Liu et al.[86], they developed multiscale dense U-Net (MD U-Net) architecture to correct drift artifacts.

An interesting application of AI is the creation of scientific chatbots to “talk” with the experiments. In this context, de Curtò et al.[87] used a methodology for the interpretation of x-ray photoelectron spectroscopy (XPS) spectra using natural language processing with LLMs. Typically used for text interpretation and generation, LLMs were adapted to reinterpret spectra and provide accessible descriptions for users, including those without specialized knowledge. Trained on synthetic data simulating complex XPS spectra, this approach effectively extracts meaningful features and simplifies the interpretation process. The integration of an LLM-informed strategy into the TANGO control[88] system at CELLS ALBA Synchrotron marks a significant advancement in synchrotron data analysis. It demonstrates the potential of AI and ML to enhance scientific research, enabling more accurate, efficient, and accessible experimentation and data analysis.

Applied-AI in real-time applications in synchrotron experiments

Real-time AI is transforming synchrotron experiments by enabling instant data analysis, allowing for the adjustment of experimental parameters. This optimization is essential for improving experimental conditions and ensuring more reliable results.[89] That is very useful in studies where samples can change rapidly of degrade during the experiments. AI can also help to enhance the temporal resolution of X-ray computed tomography (CT) scans, allowing for the analysis of dynamic behaviours of materials under varying conditions. Graas et al.[90] introduced a "just-in-time learning" DL framework for real-time X-ray CT, integrating lightweight CNNs for real-time image denoising using self-supervised Noise2Inverse learning.[81] Implemented within the RECAST3D pipeline[91], the final algorithm is able to reconstruct and analyses images in milliseconds while training the model online with spatio-temporally continuous data. The RECAST3D is an open access visualization software able to produce real-time reconstruction of images using traditional CT reconstruction algorithms. [91] Arzola-Villegas et al. used CNN to segment wood cell walls and air volumes in micro X-ray CT images.[92] By addressing the loss of intensity contrast between wood structure materials, this CNN method overcome histogram-based segmentation methods. An average intersection over union (IoU) metric of 96% indicates high CNN segmentation accuracy. Moreover, they showed that it is possible to visualize and analyse 3D cellular structures like wood water expansion in real time. Fokin et al. used U-Net[75] architectures to improve image quality and reduce noise, facilitating the study of processes like precipitation in porous media. They demonstrated the applicability of the trained model for real-time experiment steering, particularly when future time steps are unknown, and segmentation is required. On-the-fly data segmentation was employed to identify regions of interest for high-resolution scanning while continuously scanning the sample at low resolution. (see **Figure 4d**) Wang et al.[93] presented a DL-based approach to study the dynamic evolution of microstructure morphology during thin-sample solidification using synchrotron X-ray radiography. A lightweight U-Net-based CNN, combined with a self-training strategy, was developed for automatic segmentation of dendrites from radiographic images. This method efficiently handles complex and noisy datasets, enabling real-time, in-situ determination of the evolving specific interface area (SIA), which characterizes dendrite morphology.

AI algorithms can enhance real-time performance during material preparation and measurement processes (real-time feedback). Using real-time X-ray analysis, Pithan et al.[76] developed a ML-based closed-loop feedback system for autonomous thin-film and crystal growth. The system analysed X-ray reflectometry (XRR) and Bragg reflection data using 1D-CNN and multilayer perceptron (MLP). ML models can extract parameters such as thickness, density, and roughness of thin films with millisecond-scale analysis speed by embedding prior physical knowledge. Real-time feedback allows a closed-loop system able to adjust experimental conditions to stop the growth of the film at a given target thickness (**see Figure 4e**). On other hand, Rebuffi et al. described AI-driven systems capable of autonomously aligning complex X-ray mirror systems to achieve aberration-free X-ray wavefronts[94,95] and improve X-ray beam alignment.[96] There is also an reinforcement learning initiative for autonomous accelerator community (RL4AA).[97] The aim of RL4AA is to consolidate the reinforcement learning to maximize accelerator performance without the continuous human intervention. This can reduce downtime and enhance beam quality. In this context, Frank Mayet introduced an AI assistant based on large language models (LLMs) called the "General AI Assistant for Intelligent Accelerator Operations (GAIA)".[98] This assistant aims to provide operators with instant access to a vast repository of documentation, operational logs, and diagnostic information, facilitating problem resolution and optimizing operational strategies.

Applied-AI in scanning probe microscopy

Scanning probe microscopy (SPM) covers a variety of modes, each providing unique insights into the properties of surfaces at the nanoscale. Atomic force microscopy (AFM), for instance, reveals surface topography, mechanical properties, and adhesion forces, while scanning tunnelling microscopy (STM) captures electronic structures by measuring the local density of states (LDOS). More specialized

techniques, such as magnetic force microscopy (MFM) and Kelvin probe force microscopy (KPFM), allow for the investigation of magnetic domains and surface potential variations, respectively. These different modes leverage distinct interactions between the probe and sample, enabling comprehensive analysis of material properties like conductivity, elasticity, and chemical reactivity. As a result, SPM has become a versatile tool in fields ranging from materials science to biology and semiconductor research. In recent years, the integration of AI into SPM has further revolutionized nanoscale imaging and analysis.[99] A pioneering work on the application of neural networks to AFM was published by Bakucz et al.[100] in 2008. They presented a NN model to quantify the change of the AFM tip as a function of its use. Since these early efforts, AI methods have adopted in different SPM applications leading to advances in data analysis, reconstruction and tip deconvolution, automatization of measurements, speeding the scanning and improved precision in nanoscale imaging.

In automatization, Ziatdinov et al.[101] used Bayesian active learning combined with Gaussian processes (GP) and deep kernel learning (DKL) for general applications in SPM. They used PFM to demonstrate their Bayesian active learning framework. The model was trained with a few randomly chosen measurements. Then, it learned the connection between the structural features seen in the PFM data and the physical property of interest, like the hysteresis loop area. This probabilistic model predicted material properties at unmeasured points, estimating uncertainty across the spatial domain. This AI-driven approach enabled the automatic selection of measurement points, allowing for efficient and targeted data acquisition, especially when probing large datasets. On the other hand, Thomas et al. combined Gaussian regression and CNNs to explore WS_2 and Au{111} surfaces to autonomously map spectroscopic data and identify surface defects with high precision.[102] Both approaches not only reduced the time required for data acquisition but also increased reproducibility and allowed for real-time decision-making during experiments. Sotres et al. [103] combined YOLOv3 and a Siamese network to enable autonomous AFM scanning of single molecules. YOLOv3 was used as detection model to identify and locate DNA molecules in the images. It provides bounding boxes to zoom in for high-resolution imaging. Whereas the Siamese network was used to track the same molecule across multiple images. It prevents redundant imaging by distinguishing between previously imaged molecules and new targets. The system was tested to run autonomously for several days continuously, acquiring high-resolution images of single DNA molecules without user intervention.

For improve scanning speed, Kelley et al.[104] used a combination of spiral scanning, which is faster than traditional raster scanning, with AI-driven image reconstruction methods to obtain a full scanned image. A high-quality reconstructed images were obtained using a combination of compressive sensing and Gaussian process regression algorithms. The team was able to reduce data collection time by up to 5.8 times, while maintaining less than 5% error in the reconstructed images. Kim et al used CNNs to transform low-resolution (LR) AFM images into high-resolution (HR) ones and accelerating acquisition. They used CNNs with a Residual Nonlocal Attention Network (RNAN) architecture. The scanning time of the LR images was approximately one-tenth of the high-resolution (HR) one.

For the tip quality, Rashidi and Wolkov[105] applied several AI algorithms to detect when the STM tip quality deteriorates by analysing surface images. The CNN was trained on images of surface features to distinguish between a sharp tip and a degraded one, often referred to as a "double tip." When the system detects a degraded tip, it initiates an in-situ conditioning routine to restore the sharpness of the tip. In the same line, Krull et al.[106] used CNN classifiers to assess the STM tip conditions, search for good sample regions and check the quality of the images.

Another important topic in SPM is the image reconstruction. The mathematical foundations are well known[107]. But the implementation of these algorithms on real images faces challenges due to unavoidable experimental noise and tip imperfections. In this context, Bonagiri et al.[108] showed an encoder-decoder algorithm to deconvolute the AFM tip in topographic images and recover the intrinsic surface profile. The algorithm was trained using synthetic tip-convoluted images as input, and deconvoluted images as output. Kocur et al. [109] use CNNs with ResU-Net architecture to correct AFM

image artifacts. The model was trained using just synthetic data that mimic real sample structures and adding simulated artifacts and noise, reflecting common AFM image distortions like tip dilation and shadow artifacts. More advanced reconstruction was carried out by Chen et al.[110] they combined multi-view AFM images with NNs to obtain 3D reconstruction of the samples. The multi-view consisted into capture several AFM images by rotating the sample and captured multiple images from different angles. They then used a NNs to analyse and combine these multiple images, eliminating the distortions and creating a highly accurate 3D model of the surface sample. **Figure 5** shows a summary of Chen's works highlighting all the steps used to obtain 3D reconstructed images.

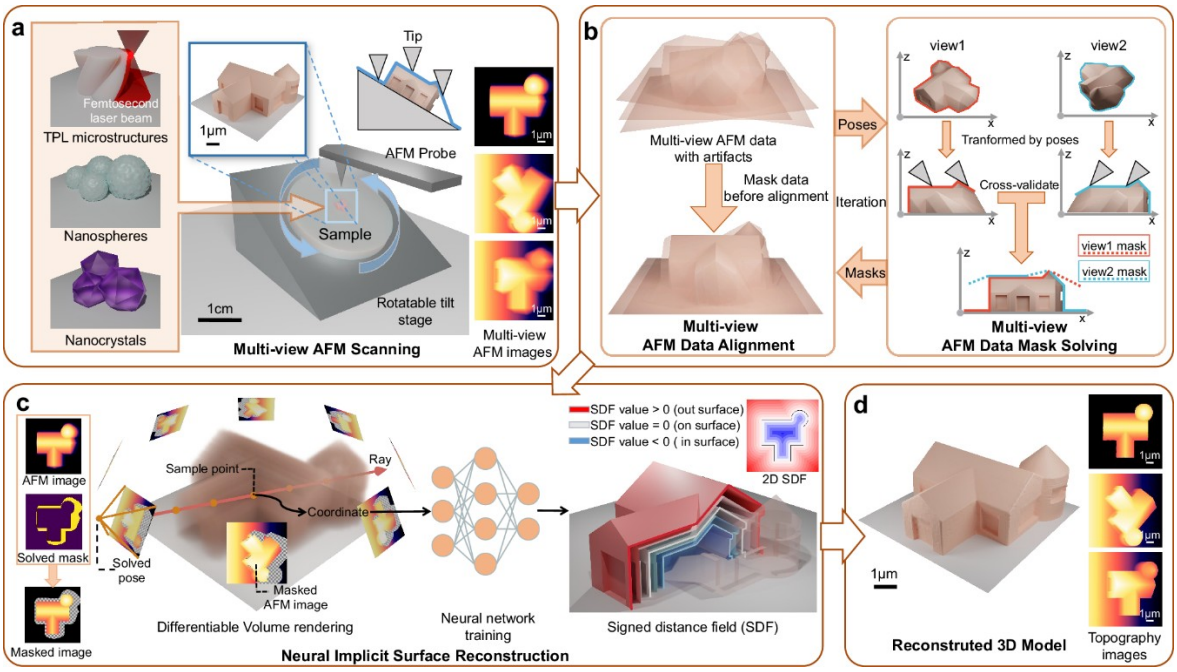


Figure 5 Application of AI in multi-view AFM images to obtain 3D reconstruction. (a) Sample fabrication and multi-view AFM scanning. (b) Multiview raw data (top) and data treatment (bottom) (c) schematic representation of the data treatment (mask) and (d) NNs analysis to reconstruct the 3D image. Reproduced with permission [110].

Finally, a very time-consuming task is the image classification. In this regard, Sokolov[111] presented a very detailed perspective article that explored the integration of ML techniques with AFM image analysis to enhance classification and characterization of sample surfaces. This perspective focused on non-deep-learning methods like decision trees and regression algorithms, which are more suitable for the limited datasets typical of AFM. The proposed ML framework enables efficient and statistically robust classification of sample surfaces by leveraging the multidimensional data captured by AFM, such as height, adhesion, and mechanical properties. A key example includes identifying cell phenotypes in cancer research, demonstrating the versatility of the approach. The observations in this perspective, particularly the methodologies for analysing biological cell surfaces, can be effectively translated to materials science for characterizing surface properties, textures, and structural variations in a wide range of material systems. Carracedo-Cosme et al.[112] used CNN and VAE networks to classify complex molecules and create synthetic AFM images, respectively. The authors created a theoretical AFM image dataset (SPMTH-60) by simulating constant-height AFM images. They used a set of 60 quasi-planar organic molecules. Then later, a synthetic dataset was generated using a VAE, which produced images closely resembling experimental AFM data. The inclusion of these VAE-generated images to the SPMTH-60 significantly enhanced the ability of the model to classify real AFM images. The same group also demonstrated the integration of high-resolution AFM (HR-AFM) with DL techniques for the molecular identification of organic compounds.[113] They used a combination of

CNNs and RNNs to translate HR-AFM images into the International Union of Pure and Applied Chemistry (IUPAC) names of molecules. The framework involves a two-step process: the first network identifies molecular attributes (e.g., functional groups), and the second network predicts the IUPAC nomenclature. Training was conducted using a massive dataset of simulated AFM images (QUAM-AFM) to ensure the robustness of the model. QUAM-AFM dataset consists of 165 million simulated AFM images generated from 686 000 molecules. These images were created using 240 different combinations of operational parameters, such as tip-sample distance and cantilever oscillation amplitude, resulting in a rich and diverse training dataset for the deep learning models. The model achieves significant accuracy, with 95% precision in attribute detection and exact molecular name predictions in 43% of cases, outperforming traditional spectroscopic methods for structural and compositional analysis. This innovation expands the capabilities of AFM, making it a powerful tool for applications in nanotechnology, catalysis, and molecular synthesis. Farley et al. [114] used CNNs to improve image segmentation of AFM images. They studied two-dimensional assemblies of gold nanoparticles formed through solvent dewetting on silicon substrates.

Applied-AI in electron microscopy

The integration of AI into (scanning) transmission electron microscopy ((S)TEM) has advanced image analysis, reduce acquisition time, accelerate the processing speed, hardware synchronization, enhancing nanocharacterization.[115–118] AI models based on Deep CNNs (e.g. Noise2Inverse[81], Noise2Void[119]), encoder-decoder[120,121] and GANs (e.g., Noise2Atom[122], CycleGAN[123,124]) architectures, have improved denoising, segmentation, and classification, even in the absence of training datasets. Among these architectures GANs represented to most popular approach, reaching beyond denoising and extending even to training dataset generation for supervised approaches (see **Figure 6**a). Supervised data analysis routines have demonstrated robustness and versability to adapt to multiple applications such as phase reconstruction from single micrographs, atom identification and tomogram reconstruction.[125–128] The integration of AI into electron microscopy (EM) has transformed data acquisition by reducing electron dose and minimizing acquisition time.[129,130] The use of compressed sensing algorithms and non-rectangular scanning paths has enabled gentler imaging of electron-beam-sensitive materials, with the potential to become a standard low-dose strategy.[131,132]

Moreover, recent advances in 4D-STEM and ptychography have opened new routes for high-resolution imaging and phase object reconstruction.[133–136] Unsupervised and DL models have been used to reconstruct phase images from 4D-STEM data to gain atomic-scale nanomaterial structural information (see **Figure 6b**).[137–139] Now, high-dimensional data includes tomography reconstructions of complex spectroscopy cubes (energy dispersive X-ray (EDX) and electron energy loss spectroscopy (EELS)) and 4D-STEM tomography, facilitated by big data handling and compressed sensing techniques (see **Figure 6c**).[140,141] As AI evolves, its integration with EM is expected to drive more sophisticated, high-dimensional analysis, such as 4D-STEM tomography combined with spectroscopy. AI has shown promise in improving 3D reconstructions, particularly for low-Z elements under low-dose conditions.[142] Algorithms like GANs and variational autoencoders, have the potential to replace traditional models in tasks like denoising and spectral unmixing.[143,144] These algorithms are self-reliant, requiring only the data being analysed, and are poised to become key tools for data engineering and training (see **Figure 6d**).[145,146]

AI has automated key aspects of EM reducing manual intervention and increasing throughput. Techniques like adaptive learning and reinforcement learning have optimized sparse scan paths and data collection.[147,148] Recent advances aim to fully control microscope optics with AI, treating lenses as trainable parameters to adapt to experimental needs and integrate custom components.[149–151] Automation in data acquisition and analysis is essential in EM[152,153], but the "human-in-the-loop" approach has gained prominence, combining AI with human expertise to improve tasks like semantic segmentation and dimensionality reduction, enhancing accuracy and reliability.[154]

Involving a trained microscopist or material scientist introduces valuable quality checkpoints in the process. First, modular routines can combine automated and manual steps, reducing the need for time-consuming dataset generation while optimizing the balance between model preparation and experiment time. Second, even in fully automated workflows, human oversight ensures the validation of intermediate AI outputs, leading to more reliable, artifact-free results.[155,156] Lastly, human involvement adds a control layer, aligning AI deployment with ALTAI principles[157], enhancing safety, and minimizing risks. The synergy between AI and EM has not only expanded existing capabilities but also unlocked new possibilities in materials characterization, driving innovations across fields such as materials science and structural biology while maximizing experimental throughput.

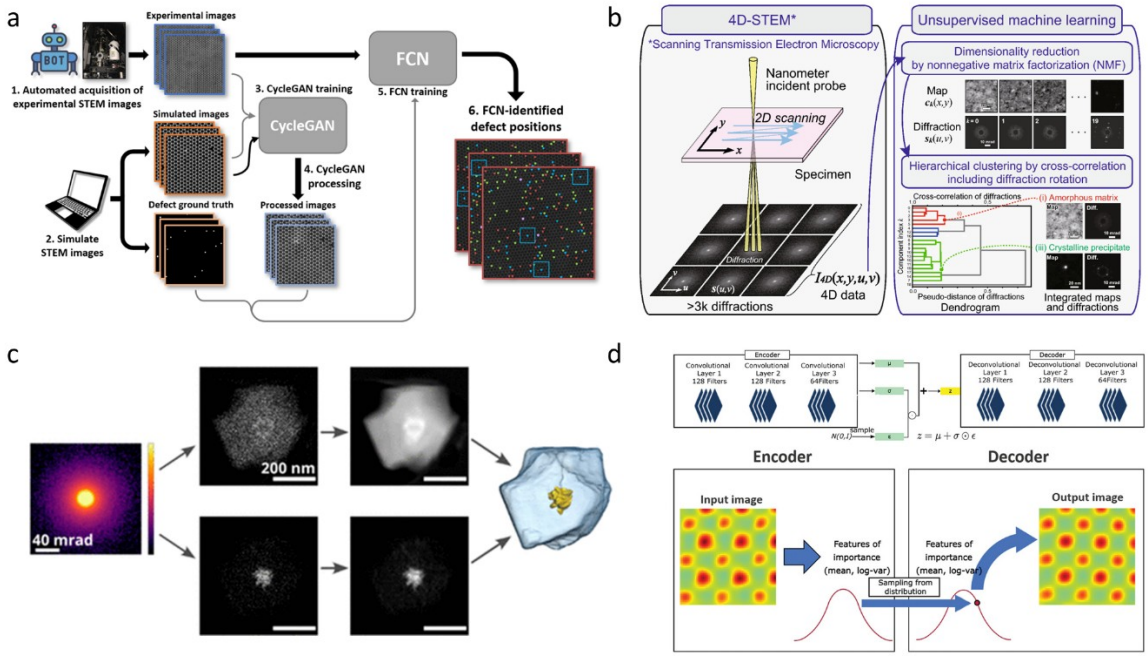


Figure 6 (a) Generative adversarial networks can be used to generate training datasets to train robust deep learning models such as convolutional neural networks for single-atom defects detection. Reprinted with permission from Ref [123]. (b) Unsupervised machine learning is a flexible choice for handling high-dimensional data such as 4D scanning transmission electron microscopy (STEM) datasets. Among others, these models can be used first to reduce dimensionality and then to cluster the outputs into differences in crystallinity within the sample. Reprinted with permission from Ref [137]. (c) Data-driven approaches to handle 4D STEM related 3D tomography reconstructions and general high-dimensional datasets. Reprinted with permission from Ref [140]. d) Encoder-decoder convolutional variational autoencoders are flexible tools to work in an unsupervised manner but also to detect anomalies in single images after appropriate training. Reprinted with permission from Ref [145].

Applied-AI in material design and property prediction

Recent advancements in AI have improved material property prediction and design. Platforms like DeepMind's AlphaFold, originally developed for protein structure prediction[11], have been extended to predict material interactions at the atomic level.[158] Although AlphaFold is mainly applied in fields like drug discovery, proteomics, and disease understanding, it has been also used to study biodegradable polymers.[159,160] Similarly, from DeepMind labs, Merchant et al.[161] have developed graph neural networks (GNNs) for material properties prediction. They introduced GNoME (Graph Networks for Materials Exploration), a system that uses advanced graph models and iterative active learning to expand the database of stable materials. By integrating computational methods like density

functional theory (DFT) and ML, the authors predicted over 2.2 million stable inorganic crystal structures, a tenfold increase compared to previous datasets. These discoveries include new layered materials, solid electrolytes and battery materials.

Collaborative platforms such as Open Catalyst uses DL to investigate catalytic reaction mechanisms.[162,163] Founded by Meta's Fundamental AI Research (FAIR) group and Carnegie Mellon University (CMU), the platform offers a dataset comprising approximately 1.3 million molecular relaxations and over 260 million density functional theory (DFT) calculations to support machine learning (ML) training.[164] It also provides a centralized repository on the Fairchem github[165], containing datasets, models, demos, and resources for applications in materials science and quantum chemistry. Additionally, Open Catalyst offers training and evaluation tools for models capable of predicting properties such as energies, forces, positions, and stresses from arbitrary chemical structures, serving as a foundational framework for research projects.

Similar efforts were also created by FAIR and Georgia Tech to develop OpenDAC dataset focused to material discovery for direct air capture (DAC) of carbon dioxide through metal-organic frameworks (MOFs).[166] The dataset includes over 176,000 adsorption energies and approximately 38 million single-point DFT calculations for CO_2, H_2O, and their mixtures on around 8000 MOFs. This large dataset enables the exploration of adsorption properties under practical conditions, including the effects of competitive water adsorption and material defects. The several scripts and the dataset is available from Fairchem repository.[165]

More recent developments include CrystaLLM, a novel framework that uses LLMs trained on millions of Crystallographic Information Files (CIFs) to generate plausible crystal structures.[167] Unlike traditional graph-based approaches, CrystaLLM uses text-based representations and autoregressive modelling to predict and explore inorganic structures. It outperforms other ML-based generative models in benchmarks for crystal structure prediction, enabling the discovery of 102 novel materials, including three thermodynamically stable ones. CrystaLLM offers an accessible web platform at https://crystallm.com, making it a valuable tool for materials exploration and accelerating discoveries in materials science. Another LLMs-based framework was recently presented by Bran et al.[168] They introduced ChemCrow, an AI-driven platform that integrates GPT-4-based LLMs for complex chemistry tasks, including organic synthesis, drug discovery, and materials design. The system allows both chemists and non-experts to automate workflows through natural language prompts.

Others tools like MatBench Discovery (https://matbench-discovery.materialsproject.org/)[169] and JARVIS (https://jarvis.nist.gov/)[170] integrate vast material datasets and high-performance ML models to predict key properties such as elasticity, bandgap, and thermal conductivity, providing critical insights for electronic and energy materials.

Applied-AI in Multiphysics modelling: metamaterials

Metamaterials offer a unique platform for implementing compactness, high speed and lower power consumption devices. As such, there is a lot of effort being put recently into proposing them in new applications, especially in cooperation with AI. Metamaterials or metastructures are artificially engineered compounds with unique properties that are not found naturally in materials. From the Greek prefix "Meta" (meaning "after" or "beyond"), the subfield aims to push normal materials beyond than their regular features. As such, the design of metamaterials plays a big role in the final functionality. The design of metamaterials typically involves arranging of micro/ nanostructures in a precise configuration to achieve a desired response.[171,172] In general, the periodicity and geometry of these elements dictate the properties of the metamaterial, allowing for tailored control over parameters such as permittivity, permeability, and chirality for the case of electromagnetic metamaterials. Although, so far this designing has been carried out via trial-and-error, some a priori knowledge or even intuition, current develops in AI have allowed to accelerate this design beyond normal human capabilities.[173,174]

While metamaterials are predominantly associated with electromagnetic applications, their unique properties and design principles have found relevance in various other fields beyond electromagnetics. One such area is acoustics, where acoustic metamaterials (or phononic crystals) have been developed to manipulate sound waves in unconventional ways. These structures have been used to control the propagation of sound, enabling applications such as sound insulation, noise cancellation, and acoustic cloaking. By leveraging the principles of negative density and negative bulk modulus, acoustic metamaterials can bend, focus, or redirect sound waves, leading to innovations in ultrasound imaging, acoustic lenses, and vibration damping systems.[175–181]

In the realm of mechanics, mechanical metamaterials often possess unusual mechanical behaviours, such as negative Poisson's ratio, negative stiffness, or extreme lightweightness and high strength materials.[182–185] Mechanical metamaterials have shown promise in applications such as shock absorption, vibration damping, and impact resistance, with potential applications in aerospace, automotive, and protective gear industries.[186–189]

Additionally, thermal metamaterials have gained attention for their ability to control the flow of heat at the nanoscale. By manipulating the thermal conductivity and heat capacity of materials through subwavelength structuring, thermal metamaterials can achieve functionalities such as thermal cloaking, thermal rectification, and thermal camouflage.[190–195] These materials hold potential for applications in thermal management, energy harvesting, and thermal insulation, offering solutions to challenges in electronics cooling, solar energy conversion, and building insulation.[196–200] Furthermore, metamaterial-inspired concepts have been explored in fields such as optics, fluid dynamics, and even robotics.[201–204]

When it comes down to modelling metamaterials, their complexity can be averaged out by defining only some effective parameters as key ingredients of the system. However, in most of cases, the complicated physical phenomena that describes “on-matter” interaction (where the term “on” refers to photon, phonon, magnon, electron, etc.) cannot be described by a generalized theory. Furthermore, when trying to calculate general spectral responses associated with very complex geometries, which is usually the case, no analytical solution is known. So, approximations become necessary, usually based on predictions from advanced iterative calculations. These types of calculations are typically carried out via numerical methods, provided by finite-differences-time-domain (FDTD) or finite element methods (FEM) to solve Maxwell or continuum elastic equations iteratively. Other quantum effects can be also added to the calculations.

In electromagnetic metamaterials, researchers typically build a 2D extended version of some unit cell, which is then called a “metasurface”. Typical efforts are focused on obtaining a structure with tailored optical properties, such as a given reflection, absorption or transmission spectrum. Deep theoretical aspects of these metamaterials have been extensively reviewed in several papers, to which we refer the reader to the works of Kadic et al. for 3D metamaterials[205], and Holloway et al. for 2D metasurfaces.[206] A key challenge lies efficiently simulating physical laws and predict material behaviour of these metamaterials. Recently, AI has emerged as powerful tools for modelling complex relationships, expanding possibilities for prediction, design and exploration of metamaterials. AI approaches to metamaterials can be categorized in various ways, such as by network architecture, material science subfield or prediction efficiency. Here, we focus on forward and inverse design. Forward design optimizes structures based on given geometrical parameters, while inverse design models work backwards from desired outcomes, such as optical spectra, to identify optimal initial geometries. Some early examples of forward design include the use CNNs for designing a polarization-insensitive plasmonic metasurface with 90% absorption,[207] and optimizing photonic crystals to maximize the quality factor.[208]

Inverse design approached the problem from the opposite direction, i.e., after the metamaterials is constructed. It uses the geometry and the electromagnetic (acoustic, phononic, etc.) spectrum as input for the algorithm (descriptors). A large dataset of geometries and spectra, typically generated via FEM

or FDTD simulations, is used to train the model (see **Figure 7a**). The goal is to enable the algorithm to identify the optimal geometry and material for a desired property, such as specific resonances or broad-band efficiency. The predicted geometry is then fabricated, often showing strong alignment with the expected optical behaviour. Despite the absence of a unified theory for metamaterial design and the computational costs of FEM/FDTD methods, these simulations still remain the most reliable means of generating training data for AI algorithms.[209,210]

Early works on the inverse design of plasmonic[211] and chiral metamaterials[212] using AI focused on a one-to-one bidirectional mapping between structural parameters and optical response (see **Figure 7B**). However, this approach may be seen limited, as different metamaterial structures can produce identical responses, challenging the uniqueness of the mapping. Such inconsistency could be treated with a forward modelling/inverse design tandem architecture.[213] Newer probabilistic graphic models have shown improvement in the previous issues, allowing to describe any random geometry beyond fixed parameterization.[214,215] Both inverse DL[214] and a deep generative forward-inverse[215] models have successfully designed optimal metamaterials, as shown in **Figure 7c** and **d**, respectively. The first model employed a GAN trained on 6500 simulations, using a pretrained simulator to ensure inversed design. The second model uses 10000 numerically generated metamaterial patterns and spectra, represented as 2D images, introducing latent variables through variational auto-encoder (VAE) structure to probabilistically represent metamaterials designs. These approaches keep several aspects of the material composition fixed and request the algorithm to determine the optimal configuration. CNNs have also been used for inverse design, where 25000 FDTD-generated spectra were used as training for randomly assigned geometrical parameters.[216] The first application of a conditional deep convolutional generative adversarial network (cDCGAN) demonstrated the ability to generate novel designs for desired optical spectra, even without prior exposure to those spectra during training.[217] Recent advancements have introduced forward/inverse methods for characterizing chiral plasmonic metamaterials using AI. Examples include multitask joint-learning in yin-yang-shaped Au nanoparticles for biosensing, eliminating the need for auxiliary networks for stabilization.[218] An accelerated deep NN that decreases training speed and improves performance in orthogonal Au nanorods[219] and dagger-shaped Ag arrays.[220] Remarkably, similar devices made of silicon nanopillars-based metasurfaces are now used as object classifiers through NN accelerators.[221,222] Additionally, topology optimization has been incorporated into inverse design[223] to meet specific topological objectives, as shown in studies on metagratings[224] and thermal-emitter metamaterials.[225]

Inverse-design using FEM simulations and AI has also made significant progresses in other metamaterials field. Mechanical metamaterials, such as auxetic structures with a negative Poisson's ratio, have been proposed through deep learning, with honeycomb lattice arrangements identified as optimal geometry.[226] Accelerated design approaches have enabled the development of these materials.[227] Moreover, recent advances in projection micro-stereolithography and NN optimization have facilitated the testing of various uniaxial compressive stress–strain curves for 3D-printed polymeric materials (see **Figure 7e**).[228] Phononic crystals (PnCs), periodic structures that can modify sound, acoustic or elastic waves, have also benefited from inverse-design efforts. Specifically, phononic metamaterials, a subset of PnCs that are locally resonant, have seen notable progress through FEM and deep learning techniques. For example, a 1D PnC designed with NNs showed high consistency with target bandgaps (see **Figure 7f**),[229] while an image-based FEM and deep learning study mapped topological structures to predicted bandgaps in phononic crystal representative volume elements.[230]

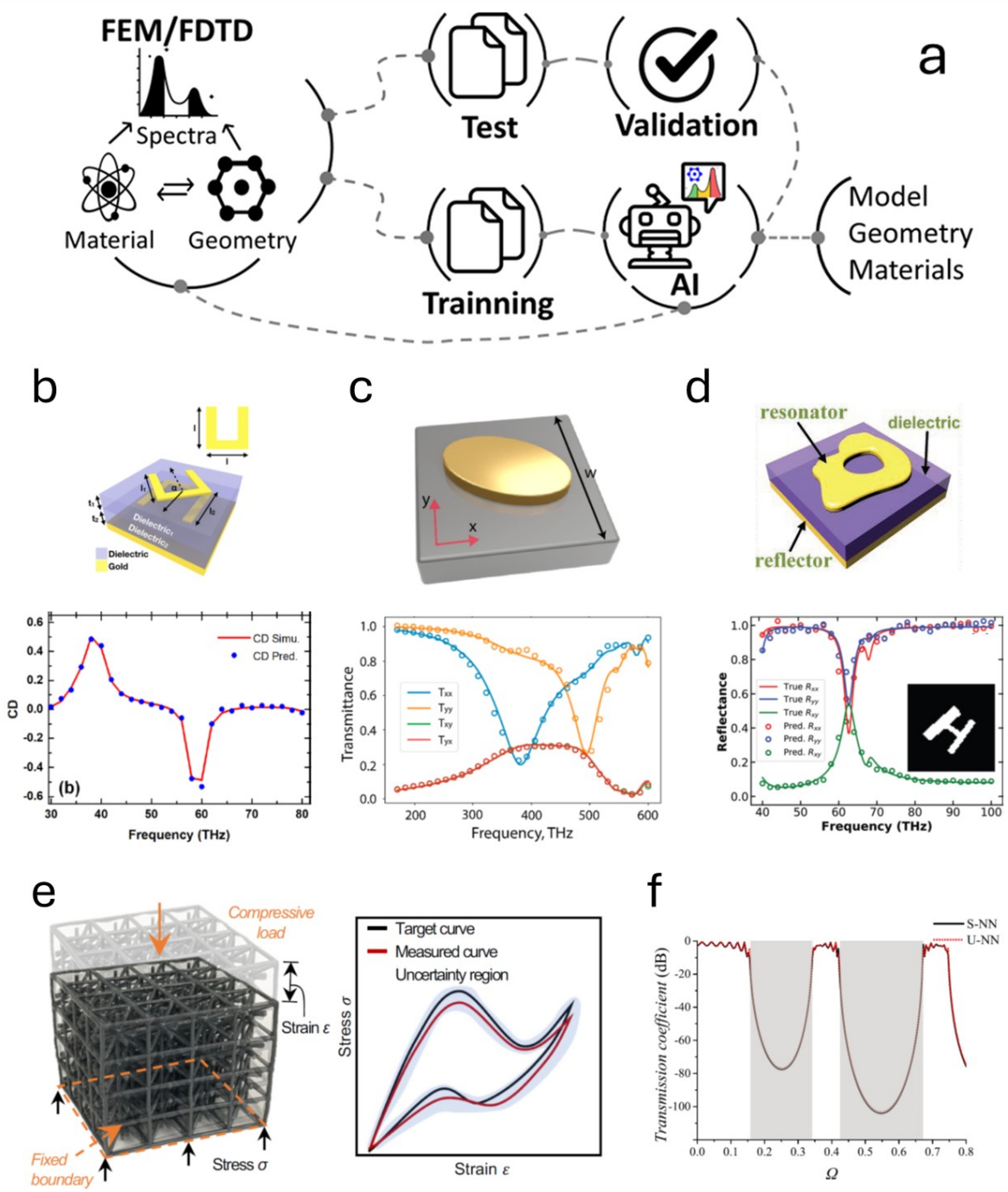


Figure 7 (a) Typical workflow for training an AI model using data generated by FEM/FDTD software. (b-d) **Electromagnetic metamaterials**: Top row shows unit cells of several cases, where the topmost unit, termed the 'resonator,' is allowed to vary in shape in (c, d) to optimize the desired spectra, while other geometrical aspects remain fixed. Bottom row compares FEM simulated spectra (lines) with AI-predicted spectra (symbols). Panel (b) is a chiral metamaterial consisting of two stacked gold split ring resonators (50 nm x 200 nm) twisted at a certain angle and separated by two spacing dielectric (n=2) layers with a continuous gold reflector at the bottom.[212] Reprinted with permission from Ref [212]. Panel (c) is a gold resonator (50 nm thick) on a glass substrate.[214] Reprinted with permission from Ref. [214] . Panel (d) is a gold resonator (50 nm thick) on a dielectric (n=2) substrate, both on top of a gold reflector below.[215] Reproduced with permission from Ref. [215]. (e) Mechanical metamaterial: A printed sample, based on the optimal design predicted by the generative ML pipeline, subjected to compressive loading. Right: A comparison of the target (black) and measured

(red) compressive stress-strain curves, with the blue shaded area indicating uncertainty due to process variability from multiple sample test.[228] Reproduced with permission from ref [228]. (f) Phononic metamaterial: Transmission coefficients of the designed 1D-PnCs with eight periods, and a two-parameter design (shear modulus ratio and mass density ratio). The gray areas represent the domains of the target bandgaps, whereas the lines represent inverse-design supervised neural network (S-NN) and unsupervised neural network (U-NN).[229] Reproduced with permission from ref [229].

Applied-AI in atomistic modelling

The equilibrium electronic properties of materials are essential for understanding their response to external electromagnetic fields, which is crucial for efficient device engineering and materials discovery. The complexity in modelling arises from the need to account for many-body interactions in systems where the number of electrons easily surpasses Avogadro's number. Current state-of-the-art methods rely heavily on DFT[231], where the energy is minimized as a function of electron density instead of solving the full Schrödinger equation, with exchange-correlation interactions incorporated self-consistently. This approach leads to the Kohn-Sham (KS) equations[232], which describe the electronic properties using single-particle KS orbitals. However, determining electronic properties using DFT remains computationally expensive due to its cubic scaling with the number of orbitals, limiting its applicability to large-scale systems. Consequently, semi-empirical methods such as the tight-binding approximation (TBA) and k·p models are often parameterized via DFT calculations and subsequently used in multiscale simulations. In TBA, atoms are modelled using a localized basis around atomic centres, which allows for the computation of the Hamiltonian matrix elements. This approach enables solving the Schrödinger or Kohn-Sham equations to obtain the spectrum and wavefunctions of the system. The diagonal elements of the Hamiltonian, or onsite energies, are related to the electronic work function, while the off-diagonal elements, known as hopping integrals, represent the energy required for an electron to move to a neighbouring orbital.[233]

TBA offers the advantage of enforcing zero hopping integrals beyond a cutoff radius, allowing efficient use of sparse matrices for simulating macroscopic materials. While local functions like Gaussian or Wannier orbitals can be used[234,235], many DFT implementations (e.g. SIESTA) employ TBA through a linear combination of atomic orbitals (LCAO) as a basis,[236] building on the foundational work of Slater and Koster for describing interactions in crystals.[237] LCAO provides a direct link between the simulation and the chemistry of the system, but it presents challenges due to the need to compute integrals over all orbitals. It increases the number of hopping terms, particularly in aperiodic systems where Bloch's theorem cannot be applied. LCAO basis sets are non-orthogonal, meaning interactions between two atoms depend not only on their pairwise relationship but also on their surrounding environment. To a first approximation, this environmental dependence can be accounted for by the inclusion of three-centre integrals, a procedure that is computationally expensive. However, recent advancements have addressed this challenge. Garrity and Choudhary integrated both two-body and three-body effective interaction terms into the TBA using a combination of high-throughput calculations and ML.[238] Their model captured local atomic environments in the Hamiltonian matrix elements, making it applicable across diverse crystal structures. Fitted to a large DFT database and improved through active learning, this approach demonstrated that parameters from training sets could be applied to novel crystal structures with high accuracy and minimal computational cost, emphasizing the critical role of atomic environments in electronic modelling.

Artificial NNs have emerged as a powerful tool in to aid the simulations, bridging the gap between the accuracy of DFT and the efficiency of TBA. In this sense, with a proper dataset, NNs enable chemically- and structure-aware parameterizations, making it possible to infer material properties from chemical formulas or atomic structures,[239] opening a door for modelling previously intractable disordered systems such as polycrystalline, amorphous materials, quasicrystals, and moiré structures, where

spatial orientation and interactions deviate from crystalline symmetry, rendering standard DFT methods computationally infeasible.[240,241]

GNNs are designed to operate on graph-structure data, where nodes correspond to entities (e.g., atoms) and edges represent relationships (e.g., bonds). GNNs iteratively update node features based on its neighbours, capturing both local and global structures as it is shown in **Figure 8**. This message-passing mechanism allows GNNs to learn complex patterns that are not easily captured by traditional models, making them particularly useful for applications like predicting material properties, molecular dynamics, and other scenarios where the data is inherently graph-based.[242,243] In material science, GNNs can model the intricate atomic interactions within crystals, capturing the essential symmetries and handling disordered systems like amorphous or polycrystalline systems. Early models, such as the Crystal Graph Convolutional Neural Network (CGCNN),[244] represent atoms as nodes and bond as edges. CGCNN uses convolutional layers to aggregate information from the neighbouring atoms, learning atomic features and interaction directly from the graph. This approach eliminates the need of manually engineered features and enables accurate prediction of materials properties, such as formation energy and band gaps. Although CGCNN achieved high accuracy in predicting these properties, its dependence on interatomic distances limits its capacity to fully capture many-body interactions, which are essential for modelling properties sensitive to bond angles and local geometric distortions. To address some of these limitations, new models improved molecular modelling by ensuring rotational and translational invariance. For example, SchNet model map the atomic positions into a learnable continuous function[245] that take the exact radial distances between atoms and apply continuous convolutions over these distances, rather than categorizing them into predefined steps. This continuous filtering allows the model to capture fine-grained variations in atomic positions and ensures that even small changes in geometry, such as slight changes in bond lengths or angles, are accounted. This innovation made SchNet particularly effective in handling molecular dynamics and non-equilibrium configurations. MEGNet, on the other hand, extended the GNN framework to both molecules and crystalline materials by incorporating global state variables, such as temperature and pressure, into its architecture. This allowed MEGNet to predict temperature-dependent properties more accurately across a broader range of materials.[246]

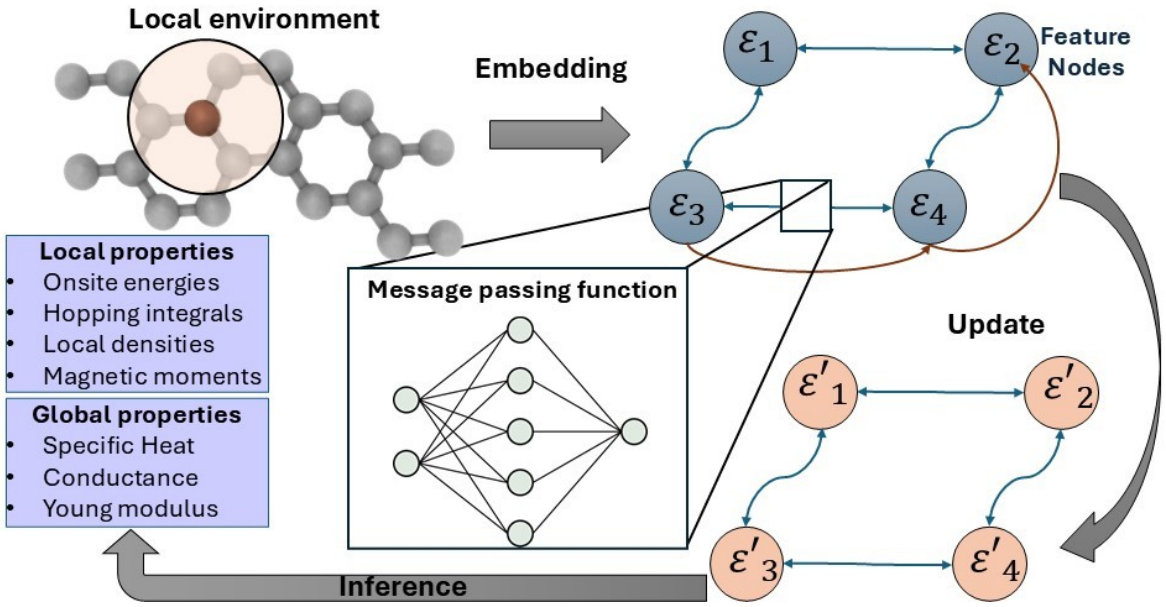


Figure 8 Scheme of the determination of local and global crystal properties using Graph Neural network. The crystal is first embedded into feature nodes, which contain the local properties. In the simplest approximation these nodes could be just the atomic information, but on more complex architecture it can also include neighbouring data and bonding information. These nodes are connected to other locally through their edges. The message passing function is in charge to collect

all local features and update the node feature. In this way, all nodes adjust themselves while considering their surroundings. Then, the graph could be used to infer both global and local properties of the crystal.

Another challenge for GNNs in materials science is their initial inability to distinguish between topologically equivalent systems that differ geometrically, such as staggered and eclipsed configurations of molecules. The Atomistic Line Graph Neural Network (ALIGNN) addressed this critical shortcoming by introducing a dual-graph approach that explicitly incorporates bond angles into the GNN framework.[247,248] In ALIGNN, bonds themselves are represented as nodes in a secondary graph, with edges representing bond angles. This structure allows the model to propagate angular information alongside interatomic distances, significantly enhancing its ability to predict properties sensitive to local geometries like band gaps. By combining bond angles with edge-gated graph convolution, ALIGNN sets a new benchmark for material property prediction, proving that explicitly capturing many-body interactions, including angular dependencies, is crucial for advancing GNN models in both molecular and crystalline systems.

In parallel, a new family of models called equivariant neural networks (ENNs) has emerged as a natural extension of CNNs. These models are designed to handle the symmetry properties of physical systems more effectively, particularly those that require strict adherence to rotational, translational, or reflectional symmetries. Equivariance refers to the property where the internal features of a model transform in the same way as its inputs under symmetry operations. This property is especially useful when constructing Hamiltonians for material systems, where symmetries like rotational and translational invariance are critical to model quantum interactions. Therefore, equivariant models and its predictions would be consistent with the inherent physical symmetries of the system, which is especially important for highly symmetric materials like crystals or those exhibiting spin-orbit coupling. For example, group equivariant CNNs (G-CNNs)[249] ensure that features are equivariant under transformations, leading to improved accuracy and generalization in structured data tasks. These models not only reduce computational complexity but also improve the predictive accuracy of electronic properties, especially in systems where symmetry plays a key role in their quantum mechanical behaviour.[250] An additional improved was achieved by introducing rotational and translational invariance into the model. This advancement tackled the challenge of accurately predicting Hamiltonian matrices for materials with intricate symmetries, such as molecular crystals and nanostructures. By ensuring that the Hamiltonian matrix transforms correctly under physical symmetries, QHNet[250] significantly improved the modelling of complex quantum interactions in three-dimensional space. Its application to molecular dynamics simulations demonstrated enhanced performance in both speed and accuracy, particularly in modelling materials with high symmetry.

The evolution of equivariant models has significantly advanced AI models for materials science, with each iteration addressing the limitations of earlier approaches. The introduction of N-centre interactions in models like equivariant representations for molecular Hamiltonians enabled the accurate prediction of complex quantum properties, improving the modelling of multi-atom interactions in systems such as large organic molecules and transition metal complexes.[251] Building on this, the equivariant analytical mapping of first principles Hamiltonians extended the transferability of these models by predicting Hamiltonians for a wide range of materials, including defected systems, using an atomic cluster expansion (ACE) descriptor that fully respects rotational symmetries.[252] The latest development, HamGNN, further integrated E(3)-equivariant convolutions, achieving exceptional accuracy and transferability across diverse material systems, from small organic molecules to large-scale, defected crystalline structures.[253] These advancements show the importance of symmetry-preserving mechanisms in improving the performance and generalization of AI models for quantum mechanical systems, with promising applications in discovery and quantum device design.

Applied-AI in molecular design: examples in pharmaceutical drug discovery

Traditionally the pharmaceutical and material science are areas of scientific interest that are quite far from each other with limited overlap. However, innovative approaches are reducing this gap, creating interdisciplinary frameworks that integrate concepts from molecules, materials, and medicine. Initiatives such as the *Molecules, Materials, Medicines* (M3) conference series[254] exemplify this convergence, fostering collaboration and advancing the integration of pharmaceutical and materials science. Materials science is critical for addressing physical and chemical challenges in drug development, such as improving solubility, bioavailability, and stability. Innovations like amorphous dispersions and pharmaceutical cocrystals exemplify the application of materials science in creating optimal drug formulations.[255] By integrating AI-driven technologies, which analyse massive datasets and predict material behaviours, with materials science techniques, researchers can refine drug formulation design, predict stability, and accelerate the discovery of novel therapeutic pathways.

This interdisciplinary approach not only addresses the bottlenecks of traditional drug development but also demonstrates how materials science principles extend beyond their traditional domain. The M3 concept emphasizes how materials science contributes to improving pharmaceutical development, while pharmaceutical challenges inspire advancements in materials science. This synergy between fields ensures innovative solutions that enhance drug performance and deliver better therapies to patients.[256]

The pharmaceutical industry plays a key role in improving global health by developing and producing essential medicines. Over the decades, this industry has evolved significantly from the manufacture of simple drugs to the creation of advanced therapies based on biotechnology and genetics. However, despite technological and scientific advances, the discovery and development of new drugs faces several traditional challenges that remain significant obstacles. One of the main challenges in traditional drug discovery is the high cost and time required to introduce a new drug to the market. The typical process includes several phases, from initial research and discovery to clinical trials, which seek to support the effectiveness and safety of new drug candidates. The process can take decades and cost billions, making it difficult to respond quickly to new diseases.[257]

On the other hand, the pharmaceutical industry is facing a declining success rate in drug discovery. Many molecules investigated in the initial phase fail to demonstrate efficacy or safety during clinical trials, resulting in a high failure rate. This not only increases costs, but also prolongs the time needed to find effective treatments. Despite these challenges, the industry has adopted new strategies to improve drug discovery. The use of technologies, such as AI with computational biology, is beginning to transform pharmaceutical research, enabling faster and more accurate analysis of large volumes of biological and chemical data. These innovations promise to reduce costs and time as well as increase success rates in the discovery of new drugs.[258,259]

Applied-AI in molecular design: example in Drug Discovery

AI has begun to play an important role in pharmacology, and its ability to process large amounts of biomedical data, such as genomic sequences, protein-ligand interactions, and clinical trial results, has accelerated the discovery of new therapeutic targets. ML and DL algorithms can identify complex subtle patterns, facilitating the identification of proteins, genes, or metabolic pathways involved in diseases that can be pharmacologically modulated. On this last point, one of the most significant advances has been the application of models of various AI tools to analyze massive genomic and proteomic data. These tools make it possible to explore genetic information associated with diseases, facilitating the identification of key proteins in pathogenesis.[260] The availability of large genomic databases (e.g., GenBank[261] or gnomAD[262]) has enabled ML and DL models to identify genetic variants associated with diseases by analyzing DNA sequences and revealing mutations and

polymorphisms. These models have proven effective in identifying patterns and associations that are difficult to identify using traditional techniques.[263] In addition, NNs have played a central role in the creation of the "interactome," a detailed mapping of protein-protein interactions. By analyzing these networks, the modulation of key proteins could have a significant impact on the disease.[264]

One of the most recent developments is the creation of DL models, such AlphaFold, which has transformed three-dimensional protein structure prediction.[11] This advance facilitates the understanding of protein function and identification of active sites for the design of molecules that can modulate the activity of target proteins.[260] The ability of these models to predict the relationship between genotype and phenotype has also allowed for a better understanding of pathogenesis by linking genomic data to clinical phenotypes, opening the door to the identification of more precise therapeutic targets.[265] The newest version of AlphaFold can predict small ligand positioning on proteins[266], protein-protein interaction modelling,[158,267] and nucleic acid simulation,[268] areas that have been a constant challenge in biomolecule modelling.

Another key approach to drug discovery is the integration of omics data. DL techniques have made it possible to combine data from multiple levels, such as genomics, transcriptomics, proteomics, and metabolomics, to provide a whole overview of pathological processes. This approach achieved by Tsuji et al.[269] facilitates the identification of metabolic pathways and transduction signals involved in the disease, which in turn suggests new therapeutic targets. Chi et al. described a novel approach to model biological networks using NNs and graph-based algorithms, allowing the representation and analysis of complex interactions in biological systems. This allows simulation of how changes in a specific protein affect the entire network and identification of therapeutic interventions that could modulate the disease.[264] Similarly, Kurata et al. combined advanced techniques, such as transferable and multitask learning, to improve therapeutic target identification. These approaches allow models to simultaneously learn from multiple diseases and datasets, which increases their ability to identify targets relevant in different pathological contexts.[270]

Advancements in modelling biological networks and potential therapeutic targets raise the need to develop efficient strategies to design molecules that interact with these targets in a precise and effective manner. The integration of advanced computational approaches, including both biological network simulations and rational drug design, is essential to accelerate the discovery of new therapeutics. As new therapeutic targets have been identified, it is important to design molecules with high affinity and specificity towards these targets, thus optimizing the therapeutic potential of the developed compounds.

Applied-AI in designing molecules with high affinity to specific targets

The design of molecules capable of binding with a high affinity to specific therapeutic targets is a key aspect in the development of new drugs and therapies. This process involves the challenge of exploring a vast and complex chemical space composed of an almost infinite number of possible combinations of chemical structures. Traditional drug design methods allow access to only a small fraction of this chemical universe, which complicates the identification of candidate molecules with therapeutic potential.[258] Pharmaceutical development is complicated because any chosen molecule must have biological activity and meet patient safety and efficacy standards. In this context, generative adversarial networks (GANs) are a promising approach in terms of practical application. They can accelerate drug discovery by creating new molecules that interact efficiently with target proteins and reduce complex data dimensionality. Variational autoencoders highlight key features in genomic data, helping to focus research efforts on areas with the greatest potential.[260]

Traditional methods like molecular docking and affinity calculation are computationally intensive and sometimes inaccurate due to model simplifications. These limitations can be overcome using new software that incorporates ML and NN. This helps predict the optimal ligand position and orientation at the active site of a receptor.[271] Limbu and Dakshanamurthy used experimental data to estimate the binding energies and dissociation constants using the hybrid DL method, improving prediction-result

correlation. This capability leads to more accurate affinity calculations, which are essential for an efficient drug design.[272]

In terms of application and recent advances, De Novo ligand generation has emerged. Using generative networks, it is possible to design new molecules with a high probability of binding to a specific target, thereby expanding the possibilities of drug discovery. Iterative optimization of models by continuously incorporating new experimental data refines predictions and improves efficiency over time. In addition, the integration of ML with other techniques such as MD provides a hybrid approach that combines the predictive capability of learning models with the detailed accuracy of physics-based methods, thus taking advantage of the best of both worlds.[273,274]

Applied-AI to drug formulation

AI is transforming the optimization of excipient selection in the development of pharmaceutical formulations, a process that traditionally relies on exhaustive experimentation and trial and error, at a high cost in time and resources. Advances in AI allow more efficient and accurate prediction of drug-excipient interactions, optimizing the final formulation.[275,276] These models predict how excipients affect crucial drug properties, such as solubility, stability, dissolution rate and bioavailability. One example is the use of NNs to predict drug solubility in various excipient combinations. For this approach, Damiati et al[277] used NNs to predict the solubility of indomethacin in aqueous solutions of different hydrotopes. They identified key features for its solubilization and selecting pyridoxine (vitamin B6) as the ideal excipient, accelerating the identification of optimal excipients.

AI not only selects excipients but also optimizes their concentration. AI models predict optimal concentration to achieve the desired disintegration and dissolution time, which is crucial in controlled-release formulations, where the excipient concentration directly affects the drug release rate. NNs can optimize parameters such as polymer concentration, coacervation agents, agitation, and spray rate. A multi-objective approach combining NNs, response surface methods and continuous genetic algorithms has been applied to investigate the effect of drug-lipid ratio, surfactant concentrations and drug loading capacity to produce polymer-lipid hybrid nanoparticles with the optimal loading and size.[278,279]

In oral formulations, ML models predict active pharmaceutical ingredient (API) solubility, API release rates from sustained-release matrix tablets, and the stability of amorphous solid dispersions.[280,281] They are also applied in the design of osmotic pump tablets, microemulsions, and drug-phospholipid/cyclodextrin complexes.[282] In nanoparticle development, AI has been applied to predict properties such as size, polydispersity index (PDI), and encapsulation efficiency, especially for transdermal delivery.[283] Finally, AI predicts conditions for preparing drug nanocrystals by techniques such as wet milling, high-pressure homogenization, or antisolvent precipitation.[284]

Applied-AI in predicting potential drug side effects.

The use of AI to predict drug toxicity and adverse reactions has gained significant attention. AI-driven analysis of preclinical and clinical data enables more accurate predictions of potential side effects of compounds, optimizing drug development. Tools such as SIDER (Side Effect Resource)[285] and ProTox 3.0[286] leverage large databases and ML models to identify relationships between drug structures and adverse effects, providing valuable platform for toxicity risks assessment.

DL models, such as DeepTox[287], can predict compound toxicity based on their molecular structures, reducing the time in safety evaluation. On the other hand, commercial tools like ADMET Predictor[288] offers comprehensive predictions for Absorption, Distribution, Metabolism, Excretion, and Toxicity (ADMET), improving the early-phase drug development. Collaborative projects like eTox Project

integrate data to develop in silico toxicity models, empathizing international cooperation to improve the accuracy of toxicological prediction.[289,290] Other tools such as VenomPred 2.0[291] and CSL-Tox[292], which employ advanced AI models to assess critical aspects of both short- and long-term toxicity, are innovative platforms that not only facilitate toxicological prediction, but also provide clear insights to guide researchers to make informed decisions about the need for further studies.

The integration of AI in drug design is a relatively recent development. However, the results obtained so far have been of great relevance and importance. Halicin, a novel antibiotic, was discovered in 2020 by researchers at the Massachusetts Institute of Technology (MIT) using AI.[293] This compound has been recognized for its ability to kill a wide range of antibiotic-resistant bacteria, including some of the most dangerous one identified by the World Health Organization (WHO). Halicin is particularly innovative because of its mechanism of action, which interferes with the ability of bacteria to maintain an electrochemical gradient across cell membranes, a process essential for energy production. Disruption of this gradient leads to bacterial death, and because this mechanism differs from conventional antibiotics, the likelihood that bacteria will develop resistance immediately is reduced.[294–296] This breakthrough was made possible by a DL algorithm, which was trained to predict molecules capable of inhibiting the growth of Escherichia coli. By analyzing the molecular structures of known drugs and their antibacterial properties, the AI system was able to analyze more than 100 million chemical compounds, stored in the "Drug Repurposing Hub" digital database, in just a few days. Halicin has emerged as a leading antibacterial candidate.[294]

Another prominent example is the case of baricitinib (Olumiant), developed by Benevolent AI and Eli Lilly for the treatment of COVID-19. Using AI-driven knowledge graphs, they managed to examine through huge biomedical datasets, identifying baricitinib as a potential treatment because of its anti-inflammatory and antiviral properties.[297,298] Baricitinib received emergency use authorization from the FDA for hospitalized patients with COVID-19. Later, it was included in treatment protocols worldwide, providing a crucial therapeutic option during the pandemic.

Finally, in silico Medicine used AI to design ISM001-055, a drug candidate for the treatment of idiopathic pulmonary fibrosis (IPF). This drug was designed using a generative enhancement tensor learning by reinforcement (GENTRL) model. This allowed the chemical exploration and molecular structure optimization of promising ligands to improve their affinity for DDR1 kinase, a key IPF pathogenesis protein.[299] It was one of the first AI-discovered drugs to enter Phase I clinical trials in 2021. This shows how AI can accelerate complex disease treatment development. **Figure 9** shows a summary of the synergy of AI and its multiple possibilities for drug prediction.

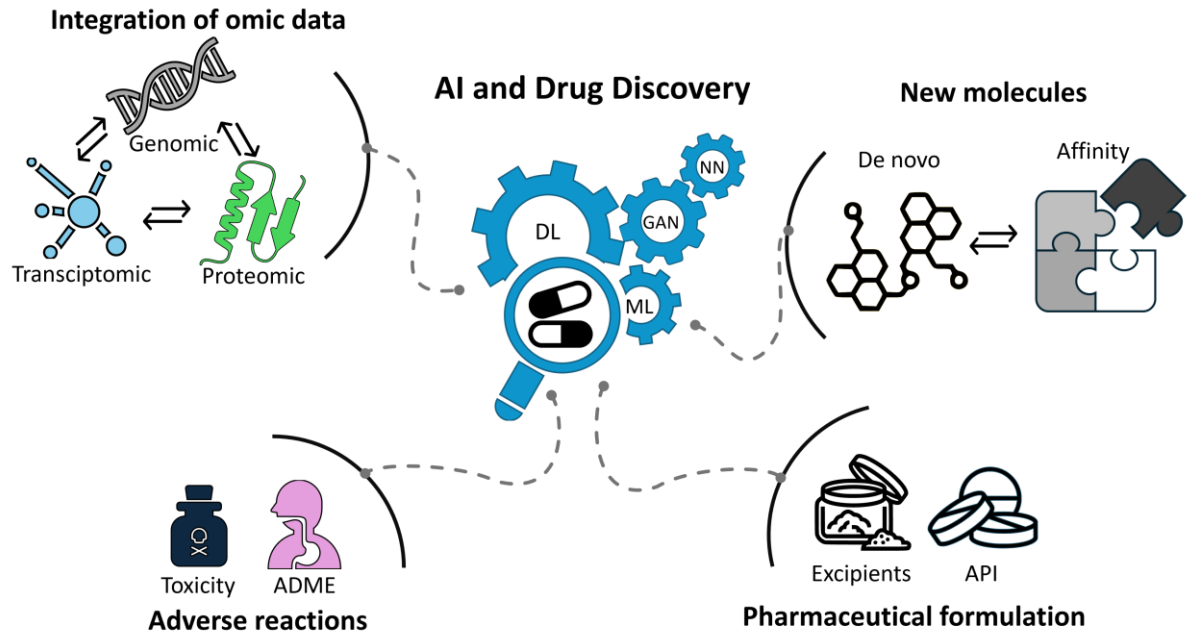


Figure 9 Scheme on the multiple uses of artificial intelligence (deep learning, machine learning, neural networks and generative adversarial networks) in drug discovery.

Expectations of AI methods

ML and AI are increasingly used in materials science. Their promise, popularity, combined with a lower barrier to entry from tutorial and frameworks. However, their widespread adoption also makes them prone to misuse. This section highlights potential challenges and corresponding solutions.

The performance of NNs is critically dependent on the training dataset. Contrary to the common perception (shaped by sensationalized news) that AI requires vast datasets processed over extensive periods using high-end GPU clusters, the actual data requirements are problem-specific. While larger datasets often improve performance, it is essential for researchers to assess the adequacy of existing datasets before assuming the necessity of acquiring additional data. In many instances, NNs can achieve satisfactory performance with smaller datasets, mitigating the significant costs and labour associated with data acquisition. Additionally, methods such as transfer learning, where a model is pre-trained on a related task before fine-tuning for the target application, can significantly reduce data requirements. Transfer learning is well-established in image classification tasks involving natural images[300] and has also been successfully applied in materials science.[301]

The statistical properties of training and test datasets are also crucial in NN development. Ideally, these datasets should be statistically indistinguishable. However, several common issues must be considered. A classic problem is overlap between training and test datasets. While direct overlap is generally avoided, subtle overlaps, such as shared subsamples from images or time-series data, may occur. Additionally, uninformative or unphysical variables, such as “IDs” or “sample names”, can inadvertently carry valuable information. To ensure the network utilizes meaningful data, researchers are advised to employ explainable AI (XAI) techniques, discussed later.

Special attention is also required when applying a network in a different condition than the training data. A common mistake is assuming that a model trained and tested on one dataset will generalize effectively. From experience, a network trained to classify centred and well-lit faces might fail when used with webcam images in less controlled environments. In a scientific context, shifts in calibration

or environmental conditions between training and application can adversely affect performance. Ensuring robustness across temporal and contextual variations is critical for reliable deployment.

Although NNs are often described as "black boxes," this label is a bit misleading. Their mathematical operations are typically straightforward, yet their behaviour can be challenging to interpret. Traditional methods, like nearest neighbours, offer greater interpretability, while techniques such as tree-based models are more transparent in simple cases but lose clarity with complex structures. Large numerical feature splits, for example, often obscure interpretability.

On the other hand, demonstrating network performance on test datasets is necessary, but often not sufficient to convince researchers of the validity of a model. Publications using NNs often devote significant effort on additional tests to convince the readers of the validity of the model. However, a simpler approach can be testing the impact of different architectures and changing the information given to the NNs during the training. In papers introducing multiple improvements, one would systematically train networks including these modifications and benchmark their performance.[302] In cases where certain data seems to be important, training can be performed with or without this data included.

XAI methods provide tools for understanding NN behaviour. For example, CNNs process input through layers of transformations that extract features. Techniques like Zeiler and Fergus's visualization method[303] allow researchers to observe the features captured at each layer. Frameworks such as LIME[304] identify regions of input (e.g., image superpixels) most influential to decisions. Similarly, gradient-based methods[305] highlight important input regions. Comparing these regions with physical expectations enhances confidence in the network's decision-making by demonstrating alignment with meaningful, domain-specific information.

Perspectives

The future of AI in materials science promises rapid advancements in discovery, design, and characterization. NNs have already helped design new materials, but several challenges remain. The training and maintenance of models are highly resource intensive. In a society where energy availability is a critical concern, the expanding energy footprint of AI presents a significant challenge. In February 2023, the chairman of Alphabet highlighted that interacting with a LLM may consume up to ten times the energy of a standard search.[306] These numbers aligned quite well with the analysis of SemiAnalysis, who estimated that ChatGPT was processing 195 million daily requests and consuming approximately 564 MWh daily or 2.9 Wh per interaction[307] compared with a standard Google search that uses ~0.3 Wh per search.[308] The substantial energy demands of AI models, such as those used for material discovery, highlight the need for integrating sustainable practices and energy-efficient algorithms. The adoption of new architectures or NNs can reduce the demand of energy. Spiking Neural Networks, the so-called third generation of NNs, are known for their efficiency in processing spatiotemporal information with minimal energy consumption.[309] The integration of SNNs with Neural Architecture Search (NAS) can further enhance their efficiency and accuracy, while preserving low latency and energy requirements.[310]

In spectroscopic data analysis, the developing models adaptable to various techniques without extensive retraining remains a priority. Interoperability issues arise as models designed for specific techniques, such as Raman spectroscopy, may not directly apply to others like XRD, UV-Vis, or NMR without modifications. Variability in data quality, influenced by instrument sensitivity, environmental conditions, and sample properties, also requires robust models capable of accurate analysis under different conditions. Meta-learning algorithms offer a solution, enabling AI models to adapt quickly to new spectroscopic methods, enhancing versatility and reducing retraining needs. On other hand, the introduction of XAI in AI-guided spectroscopy is crucial because it transforms complex, opaque ML models into transparent systems that provide interpretable insights. XAI enables researchers to determine key spectral bands or peaks influencing outcomes, making AI tools more reliable and

fostering their adoption in other fields. This interpretability bridges the gap between advanced AI techniques and domain expertise, ensuring that AI models are both effective and trustworthy.

Similar challenges faced in spectroscopy affect AI applications in EM, where conditions like beam intensity, magnification, and sample composition vary significantly. Models trained under specific conditions may struggle to generalize, suggesting that training models on shared datasets across laboratories could enhance robustness and applicability. Moreover, standardizing AI tools across synchrotron facilities, which uses diverse hardware and software, is crucial for consistent data processing. Global partnerships projects and modular AI platforms tailored for different synchrotron technologies could foster international collaboration and efficiency. The In-situ Correlative Facility for Advanced Energy Materials (InCAEM) is a pioneering initiative in this direction. Its integration with ALBA Synchrotron will enable combined (S)TEM, SPM, and synchrotron techniques to provide atomic-scale insights under real working conditions. For such endeavour, the development of AI-based advanced algorithms capable of processing and correlating massive multi-modal datasets in real-time will be essential for improving the speed and accuracy of scientific discovery. That will also require the harmonisation of correlative data from different experimental techniques, ensuring consistency and accuracy when combining datasets from microscopy and synchrotron sources.

In material design and atomistic modelling, intensive simulations, such as FEM, FDTD, and DFT, are computationally expensive, slowing down the iterative process of material design. The physical assumptions and approximations in these models can also introduce errors, making it challenging to achieve high precision without compromising speed. The reliance on complex "*black-box"* models in metamaterial inverse design further complicates understanding and trust in AI-generated outcomes. Creating explainable AI models that provide the justification for design choices would improve the acceptance and confidence of the field. In this sense, it is important to remark again the early adoption of XAI tools.

The role of AI in drug discovery has shown promise, but integrating these tools into pharmaceutical processes introduces complexities. Rigorous validation is needed for regulatory approval, bridging the gap between AI predictions and compliance standards. Bias in AI drug discovery models, particularly when training data lacks diversity, is another concern. Collaborative platforms involving pharmaceutical companies, AI developers, and regulatory bodies could streamline AI integration. Diversifying training datasets with genomic data from varied populations can also reduce biases and enhance model robustness.

An additional critical aspect in the integration of AI into materials science and drug discovery is the need for experimentalists to adopt these AI tools. While the development and optimization of AI models are often driven by computer scientists and engineers, it is essential that experimental researchers (who generate and work with the data) are well-trained in the use these tools. Their expertise in experimental methodologies combined with an understanding of AI applications will enable them to interpret AI outputs accurately and integrate these tools into their research. Training programs and interdisciplinary workshops aimed at equipping experimentalists with the skills to use AI platforms are crucial for maximizing the potential of this technology. By fostering collaboration between AI developers and experimental scientists, the field can ensure that AI tools are not only accessible but also practically implemented to achieve more precise and efficient research outcomes.

In summary, the future of AI in materials science and drug discovery face some technical, ethical, and infrastructural challenges. Cross-disciplinary integration is essential, requiring collaboration among computer scientists, physicists, chemists, and engineers to align goals and methodologies. Ethical considerations, such as data ownership, intellectual property, and transparency, also demand attention. Establishing global consortiums and interdisciplinary research hubs can bridge these gaps, while implementing ethical governance frameworks ensures responsible development. Addressing these challenges and leveraging collaborative approaches are critical for AI's advancement and

transformative impact in these fields. Facilities like InCAEM will be key in addressing these challenges, fostering AI-driven innovations in sustainable material design and energy applications.

Author contributions

E.C.-A. and A.C.A. designed, reviewed, and edited the entire manuscript. E.C.-A. wrote and edited spectroscopic, synchrotron and SPM sections. M.B.E. wrote the introduction, expectations and perspective sections; A.C.A. wrote and edited the pharmaceutical and help with synchrotrons sections. J.H.G wrote the DFT section. M.B.M. wrote the EM section. O.A.O. wrote the metamaterials section. J.A. wrote the EM section. A.M. supervised the whole project and review the spectroscopy, synchrotron and SPM section. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

ICN2 is funded by the CERCA programme/Generalitat de Catalunya and by the Severo Ochoa programme of the Spanish Ministry of Economy, Industry and Competitiveness (MINECO, grant no. SEV-2017-0706). We acknowledge the financial support from MCIN with funding from NextGenerationEU (PRTR-C17.I1) and Generalitat de Catalunya. A.C.A acknowledges support from Chilean National Fund for Scientific and Technological Development (FONDECYT) N° 11200620. O.A.-O. acknowledges the United States-Israel Binational Science Foundation BSF (Grant No. 2018050), and the NQPI at Ohio University. M.B.M and J.A. acknowledge support from CSIC Interdisciplinary Thematic Platform (PTI+) on Quantum Technologies (PTI-QTEP+), European Commission– NextGenerationEU (Regulation EU 2020/2094), through CSIC's Quantum Technologies Platform (QTEP). Part of the present work has been performed in the framework of Universitat Autònoma de Barcelona Materials Science PhD program. M.B.M acknowledges support from SUR Generalitat de Catalunya and the EU Social Fund; project ref. 2020 FI 00103. This study was supported by EU HORIZON INFRA TECH 2022 project IMPRESS (Ref.: 101094299). ICN2 acknowledges funding from Grant IU16-014206 (METCAM FIB) funded by the European Union through the European Regional Development Fund (ERDF), with the support of the Ministry of Research and Universities, Generalitat de Catalunya. ICN2 is founding member of e-DREAM.[311]

References

[1] S. Haykin, *Neural Networks and Learning Machines*, Pearson Education, Upper Saddle River, **2011**.

[2] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, Li Fei-Fei, in *2009 IEEE Conf. Comput. Vis. Pattern Recognit.*, IEEE, **2009**, pp. 248–255.

[3] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, L. Fei-Fei, *Int. J. Comput. Vis.* **2015**, *115*, 211.

[4] A. Krizhevsky, I. Sutskever, G. E. Hinton, *Commun. ACM* **2017**, *60*, 84.

[5] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, in *Adv. Neural Inf. Process. Syst.* (Eds.: Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, K. Q. Weinberger), Curran Associates, Inc., **2014**.

[6] A. Brock, J. Donahue, K. Simonyan, in *ICLR 2019 Conf.*, **2019**.

[7] J. Sohl-Dickstein, E. A. Weiss, N. Maheswaranathan, S. Ganguli, in *Proc. 32nd Int. Conf. Int. Conf. Mach. Learn. - Vol. 37*, JMLR.Org, **2015**, pp. 2256–2265.

[8] J. Ho, A. Jain, P. Abbeel, in *Proc. 34th Int. Conf. Neural Inf. Process. Syst.*, Curran Associates Inc., Red Hook, NY, USA, **2020**.

[9] A. Martín, A. Ashish, B. Paul, B. Eugene, C. Zhifeng, C. Craig, C. Greg S., D. Andy, D. Jeffrey, D. Matthieu, G. Sanjay, G. Ian, H. Andrew, I. Geoffrey, I. Michael, Y. Jia, J. Rafal, K. Lukasz, K.

Manjunath, L. Josh, M. Dandelion, M. Rajat, M. Sherry, M. Derek, O. Chris, S. Mike, S. Jonathon, S. Benoit, S. Ilya, T. Kunal, T. Paul, V. Vincent, V. Vijay, V. Fernanda, V. Oriol, W. Pete, W. Martin, W. Martin, Y. Yuan, Z. Xiaoqiang, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, **2015**.

[10] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, S. Chintala, in *Proc. 33rd Int. Conf. Neural Inf. Process. Syst.*, Curran Associates Inc., Red Hook, NY, USA, **2019**.

[11] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, *Nature* **2021**, *596*, 583.

[12] J. Rodríguez, I. Rodríguez-Rodríguez, W. L. Woo, *WIREs Data Min. Knowl. Discov.* **2022**, *12*, DOI 10.1002/widm.1476.

[13] C. J. Fluke, C. Jacobs, *WIREs Data Min. Knowl. Discov.* **2020**, *10*, DOI 10.1002/widm.1349.

[14] H.-L. Xu, T.-T. Gong, F.-H. Liu, H.-Y. Chen, Q. Xiao, Y. Hou, Y. Huang, H.-Z. Sun, Y. Shi, S. Gao, Y. Lou, Q. Chang, Y.-H. Zhao, Q.-L. Gao, Q.-J. Wu, *eClinicalMedicine* **2022**, *53*, 101662.

[15] S. Bhattacharya, R. K. Mahato, S. Singh, G. K. Bhatti, S. S. Mastana, J. S. Bhatti, *Life Sci.* **2023**, *332*, 122110.

[16] "Web of Science.," can be found under www.webofscience.com, **n.d.**

[17] F. Kibrete, T. Trzepieciński, H. S. Gebremedhen, D. E. Woldemichael, *J. Compos. Sci.* **2023**, *7*, 364.

[18] N. J. Szymanski, B. Rendy, Y. Fei, R. E. Kumar, T. He, D. Milsted, M. J. McDermott, M. Gallant, E. D. Cubuk, A. Merchant, H. Kim, A. Jain, C. J. Bartel, K. Persson, Y. Zeng, G. Ceder, *Nature* **2023**, *624*, 86.

[19] J. S. Huang, J. X. Liew, A. S. Ademiloye, K. M. Liew, *Arch. Comput. Methods Eng.* **2021**, *28*, 3399.

[20] R. Gurnani, S. Shukla, D. Kamal, C. Wu, J. Hao, C. Kuenneth, P. Aklujkar, A. Khomane, R. Daniels, A. A. Deshmukh, Y. Cao, G. Sotzing, R. Ramprasad, *Nat. Commun.* **2024**, *15*, 6107.

[21] A. Ravichandran, S. Honrao, S. Xie, E. Fonseca, J. W. Lawson, *J. Phys. Chem. Lett.* **2024**, *15*, 121.

[22] A. N. Abdalla, M. S. Nazir, H. Tao, S. Cao, R. Ji, M. Jiang, L. Yao, *J. Energy Storage* **2021**, *40*, 102811.

[23] B. Roter, S. V. Dordevic, *Phys. C Supercond. its Appl.* **2020**, *575*, 1353689.

[24] X. Jia, A. Aziz, Y. Hashimoto, H. Li, *Sci. China Mater.* **2024**, *67*, 1173.

[25] H. Tiryaki, A. Yusuf, S. Ballikaya, *Energy* **2024**, *292*, 130597.

[26] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, *npj Comput. Mater.* **2017**, *3*, 54.

[27] V. Tshitoyan, J. Dagdelen, L. Weston, A. Dunn, Z. Rong, O. Kononova, K. A. Persson, G. Ceder, A. Jain, *Nature* **2019**, *571*, 95.

[28] ALBA, "In Situ Correlative Facility for Advanced Energy Materials: InCAEM," can be found under https://www.cells.es/en/instruments/electron-microscopes/incaem-project, **2023**.

[29] R. J. Angel, M. Murri, B. Mihailova, M. Alvaro, *Zeitschrift für Krist. - Cryst. Mater.* **2019**, *234*, 129.

[30] S. Sandell, E. Chávez-Ángel, A. El Sachat, J. He, C. M. Sotomayor Torres, J. Maire, *J. Appl. Phys.* **2020**, *128*, 131101.

[31] S. Palleschi, D. Mastrippolito, P. Benassi, M. Nardone, L. Ottaviano, *Appl. Surf. Sci.* **2021**, *561*, 149691.

[32] J. A. García-Merino, R. Villarroel, E. Chávez-Ángel, S. A. Hevia, *Opt. Laser Technol.* **2024**, *178*, 111255.

[33] D. B. Schuepfer, F. Badaczewski, J. M. Guerra-Castro, D. M. Hofmann, C. Heiliger, B. Smarsly, P. J. Klar, *Carbon N. Y.* **2020**, *161*, 359.

[34] Z. Han, X. Yang, S. E. Sullivan, T. Feng, L. Shi, W. Li, X. Ruan, *Phys. Rev. Lett.* **2022**, *128*, 045901.

[35] X. Chu, Y. Huang, Y.-H. Yun, X. Bian, *Chemometric Methods in Analytical Spectroscopy Technology*, Springer Nature Singapore, Singapore, **2022**.

[36] R. G. Brereton, *J. Chemom.* **2014**, *28*, 749.

[37] M. Greenacre, P. J. F. Groenen, T. Hastie, A. I. D'Enza, A. Markos, E. Tuzhilina, *Nat. Rev. Methods Prim.* **2022**, *2*, 100.

[38] M. Greenacre, P. J. F. Groenen, T. Hastie, A. I. D'Enza, A. Markos, E. Tuzhilina, *Nat. Rev. Methods Prim.* **2023**, *3*, 22.

[39] S. Wold, K. Esbensen, P. Geladi, *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37.

[40] D. Mainka, J. Krause, L. Großmann, A. Link, L. Schulig, *J. Chem. Educ.* **2023**, *100*, 2435.

[41] A. Becht, C. Schollmayer, Y. Monakhova, U. Holzgrabe, *Anal. Bioanal. Chem.* **2021**, *413*, 3107.

[42] E. Chavez-Angel, B. Puertas, M. Kreuzer, R. Soliva Fortuny, R. C. Ng, A. Castro-Alvarez, C. M. Sotomayor Torres, *Foods* **2022**, *11*, 1304.

[43] E. J. Rifna, R. Pandiselvam, A. Kothakota, K. V. Subba Rao, M. Dwivedi, M. Kumar, R. Thirumdas, S. V. Ramesh, *Food Chem.* **2022**, *369*, 130898.

[44] A. Biancolillo, F. Marini, C. Ruckebusch, R. Vitale, *Appl. Sci.* **2020**, *10*, 6544.

[45] N. M. Ralbovsky, I. K. Lednev, *Chem. Soc. Rev.* **2020**, *49*, 7428.

[46] Z. Wang, Z. Sun, H. Yin, X. Liu, J. Wang, H. Zhao, C. H. Pang, T. Wu, S. Li, Z. Yin, X. Yu, *Adv. Mater.* **2022**, *34*, DOI 10.1002/adma.202104113.

[47] L. Pan, P. Zhang, C. Daengngam, S. Peng, M. Chongcheawchamnan, *J. Raman Spectrosc.* **2022**, *53*, 6.

[48] A. de Juan, R. Tauler, *Anal. Chim. Acta* **2021**, *1145*, 59.

[49] M. Sledzinska, P. Xiao, E. Puig Vilardell, E. Chávez Angel, M. J. Esplandiu, C. M. Sotomayor Torres, *Appl. Phys. Lett.* **2022**, *121*, DOI 10.1063/5.0130927.

[50] L. Matthies, M. T. Gebrekidan, A. S. Braeuer, R. E. Friedrich, F. Stelzle, C. Schmidt, R. Smeets, A. T. Assaf, M. Gosau, T. Rolvien, C. Knipfer, *Oral Dis.* **2024**, *30*, 2439.

[51] X. Y. Zhao, G. Y. Liu, Y. T. Sui, M. Xu, L. Tong, *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.* **2021**, *250*, 119374.

[52] M. T. Gebrekidan, C. Knipfer, A. S. Braeuer, *J. Raman Spectrosc.* **2021**, *52*, 723.

[53] J. Brandt, K. Mattsson, M. Hassellöv, *Anal. Chem.* **2021**, *93*, 16360.

[54] Y. Zeng, Z. Liu, X. Fan, X. Wang, *Microchem. J.* **2023**, *191*, 108777.

[55] J. Lim, G. Shin, D. Shin, *Anal. Chem.* **2024**, *96*, 6819.

[56] O. Guselnikova, A. Trelin, Y. Kang, P. Postnikov, M. Kobashi, A. Suzuki, L. K. Shrestha, J. Henzie, Y. Yamauchi, *Nat. Commun.* **2024**, *15*, 4351.

[57] X. Yan, Z. Cao, A. Murphy, Y. Ye, X. Wang, Y. Qiao, *Sci. Total Environ.* **2023**, *896*, 165340.

[58] J. Contreras, T. Bocklitz, *Pflügers Arch. - Eur. J. Physiol.* **2024**, DOI 10.1007/s00424-024-02997-y.

[59] X. Zhong, B. Gallagher, S. Liu, B. Kailkhura, A. Hiszpanski, T. Y.-J. Han, *npj Comput. Mater.* **2022**, *8*, 204.

[60] J. Yoo, Y. Cho, B. Jeong, S. H. Choi, K. K. Kim, S. C. Lim, S. M. Lee, J. Choo, M. S. Jeong, *Adv. Intell. Syst.* **2023**, *5*, DOI 10.1002/aisy.202200463.

[61] J. Yoo, Y. Cho, D. H. Kim, J. Kim, T. G. Lee, S. M. Lee, J. Choo, M. S. Jeong, *Nano Today* **2024**, *57*, 102366.

[62] P. Solís-Fernández, H. Ago, *ACS Appl. Nano Mater.* **2022**, *5*, 1356.

[63] T. Vincent, K. Kawahara, V. Antonov, H. Ago, O. Kazakova, *Carbon N. Y.* **2023**, *201*, 141.

[64] D. Hu, T.-F. Chung, Y. P. Chen, Y. Qi, **2024**.

[65] E. Chavez-Angel, R. C. Ng, S. Sandell, J. He, A. Castro-Alvarez, C. M. Sotomayor Torres, M. Kreuzer, *Polymers (Basel).* **2023**, *15*, 536.

[66] D. Georgiev, S. V. Pedersen, R. Xie, Á. Fernández-Galiana, M. M. Stevens, M. Barahona, *Anal. Chem.* **2024**, *96*, 8492.

[67] M. Toplak, S. T. Read, C. Sandt, F. Borondics, L. Vaccari, H. J. Byrne, T. P. Wrobel, *Cells 2021, Vol. 10, Page 2300* **2021**, *10*, 2300.

[68] M. Toplak, G. Birarda, S. Read, C. Sandt, S. M. Rosendahl, L. Vaccari, J. Demšar, F. Borondics, *Synchrotron Radiat. News* **2017**, *30*, 40.

[69] J. Demsar, T. Curk, A. Erjavec, C. Gorup, T. Hocevar, M. Milutinovic, M. Mozina, M. Polajnar, M. Toplak, A. Staric, M. Stajdohar, L. Umek, L. Zagar, J. Zbontar, M. Zitnik, B. Zupan, *J. Mach. Learn. Res.* **2013**, *14*, 2349.

[70] J. Li, X. Huang, P. Pianetta, Y. Liu, *Nat. Rev. Phys.* **2021**, *3*, 766.

[71] C. Wang, U. Steiner, A. Sepe, *Small* **2018**, *14*, DOI 10.1002/smll.201802291.

[72] Z. Zhang, C. Li, W. Wang, Z. Dong, G. Liu, Y. Dong, Y. Zhang, *Innov.* **2024**, *5*, 100539.

[73] X. Yang, M. Kahnt, D. Brückner, A. Schropp, Y. Fam, J. Becher, J.-D. Grunwaldt, T. L. Sheppard, C. G. Schroer, *J. Synchrotron Radiat.* **2020**, *27*, 486.

[74] A. A. Hendriksen, M. Bührer, L. Leone, M. Merlini, N. Vigano, D. M. Pelt, F. Marone, M. di Michiel, K. J. Batenburg, *Sci. Rep.* **2021**, *11*, 11895.

[75] M. I. Fokin, V. V. Nikitin, A. A. Duchkov, *J. Synchrotron Radiat.* **2023**, *30*, 978.

[76] L. Pithan, V. Starostin, D. Mareček, L. Petersdorf, C. Völter, V. Munteanu, M. Jankowski, O. Konovalov, A. Gerlach, A. Hinderhofer, B. Murphy, S. Kowarik, F. Schreiber, *J. Synchrotron Radiat.* **2023**, *30*, 1064.

[77] J. E. Salgado, S. Lerman, Z. Du, C. Xu, N. Abdolrahim, *npj Comput. Mater.* **2023**, *9*, 214.

[78] Y. Suzuki, H. Hino, T. Hawai, K. Saito, M. Kotsugi, K. Ono, *Sci. Rep.* **2020**, *10*, 21790.

[79] N. E. Omori, A. D. Bobitan, A. Vamvakeros, A. M. Beale, S. D. M. Jacques, *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2023**, *381*, DOI 10.1098/rsta.2022.0350.

[80] M. Zhang, S. Gu, Y. Shi, *Complex Intell. Syst.* **2022**, *8*, 5545.

[81] A. A. Hendriksen, D. M. Pelt, K. J. Batenburg, *IEEE Trans. Comput. Imaging* **2020**, *6*, 1320.

[82] Z. Liu, T. Bicer, R. Kettimuthu, D. Gursoy, F. De Carlo, I. Foster, *J. Opt. Soc. Am. A* **2020**, *37*, 422.

[83] F. De Carlo, D. Gürsoy, D. J. Ching, K. J. Batenburg, W. Ludwig, L. Mancini, F. Marone, R. Mokso, D. M. Pelt, J. Sijbers, M. Rivers, *Meas. Sci. Technol.* **2018**, *29*, 034004.

[84] X. Yang, F. De Carlo, C. Phatak, D. Gürsoy, *J. Synchrotron Radiat.* **2017**, *24*, 469.

[85] "Xlearn: An open-source Python package implementing Deep Learning for X-ray Science," can be found under https://xlearn.readthedocs.io/en/latest/, **2016**.

[86] M. Liu, Y. Han, X. Xi, L. Zhu, S. Yang, S. Tan, J. Chen, L. Li, B. Yan, *Entropy* **2022**, *24*, 967.

[87] J. de Curtò, I. de Zarzà, G. Roig, C. T. Calafate, *Signals* **2024**, *5*, 181.

[88] J.-M. . Chaize, A. . Götz, W.-D. . Klotz, J. . Meyer, P. M., T. E., in *7th Int. Conf. Accel. Large Exp. Phys. Control Syst.*, Trieste, **1999**.

[89] J. E. McClure, J. Yin, R. T. Armstrong, K. C. Maheshwari, S. Wilkinson, L. Vlcek, Y. Da Wang, M. A. Berrill, M. Rivers, **2020**, pp. 226–239.

[90] A. Graas, S. B. Coban, K. J. Batenburg, F. Lucka, *Sci. Rep.* **2023**, *13*, 20070.

[91] J.-W. Buurlage, H. Kohr, W. Jan Palenstijn, K. Joost Batenburg, *Meas. Sci. Technol.* **2018**, *29*, 064005.

[92] X. Arzola-Villegas, C. Báez, R. Lakes, D. S. Stone, J. O'Dell, P. Shevchenko, X. Xiao, F. De Carlo, J. E. Jakes, *Appl. Sci.* **2023**, *13*, 8146.

[93] N. Wang, Y. Tang, Y. Wu, Y. Zhang, Y. Dai, J. Zhang, R. Zhang, Y. Xu, B. Sun, *Mater. Charact.* **2021**, *181*, 111451.

[94] L. Rebuffi, X. Shi, R. Zhang, M. J. Highland, M. G. Frith, W. Cha, R. J. Harder, S. Kandel, M. J. Cherukara, L. Assoufid, in *Adv. Comput. Methods X-Ray Opt. VI* (Eds.: O. Chubar, T. Tanaka), SPIE, **2023**, p. 12.

[95] L. Rebuffi, X. Shi, Z. Qiao, M. J. Highland, M. G. Frith, A. Wojdyla, K. A. Goldberg, L. Assoufid, *Opt. Express* **2023**, *31*, 21264.

[96] L. Rebuffi, S. Kandel, X. Shi, R. Zhang, R. J. Harder, W. Cha, M. J. Highland, M. G. Frith, L. Assoufid, M. J. Cherukara, *Opt. Express* **2023**, *31*, 39514.

[97] A. Santamaria Garcia, A. Eichler, C. Xu, J. Kaiser, L. Scomparin, M. Schenk, S. Pochaba, S. Hirlaender, in *IPAC'24 - 15th Int. Part. Accel. Conf.*, JACoW Publishing, Nashville, TN, **2024**, pp. 1812–1815.

[98] F. Mayet, **2024**.

[99] M. A. Rahman Laskar, U. Celano, *APL Mach. Learn.* **2023**, *1*, DOI 10.1063/5.0160568.

[100] P. Bakucz, A. Yacoot, T. Dziomba, L. Koenders, R. Krüger-Sehm, *Meas. Sci. Technol.* **2008**, *19*, 065101.

[101] M. Ziatdinov, Y. Liu, K. Kelley, R. Vasudevan, S. V. Kalinin, *ACS Nano* **2022**, *16*, 13492.

[102] J. C. Thomas, A. Rossi, D. Smalley, L. Francaviglia, Z. Yu, T. Zhang, S. Kumari, J. A. Robinson, M. Terrones, M. Ishigami, E. Rotenberg, E. S. Barnard, A. Raja, E. Wong, D. F. Ogletree, M. M.

Noack, A. Weber-Bargioni, *npj Comput. Mater.* **2022**, *8*, 99.

[103] J. Sotres, H. Boyd, J. F. Gonzalez-Martinez, *Nanoscale* **2021**, *13*, 9193.

[104] K. P. Kelley, M. Ziatdinov, L. Collins, M. A. Susner, R. K. Vasudevan, N. Balke, S. V. Kalinin, S. Jesse, *Small* **2020**, *16*, DOI 10.1002/smll.202002878.

[105] M. Rashidi, R. A. Wolkow, *ACS Nano* **2018**, *12*, 5185.

[106] A. Krull, P. Hirsch, C. Rother, A. Schiffrin, C. Krull, *Commun. Phys.* **2020**, *3*, 54.

[107] J. S. Villarrubia, *J. Res. Natl. Inst. Stand. Technol.* **1997**, *102*, 425.

[108] L. K. S. Bonagiri, Z. Wang, S. Zhou, Y. Zhang, *Nano Lett.* **2024**, *24*, 2589.

[109] V. Kocur, V. Hegrová, M. Patočka, J. Neuman, A. Herout, *Ultramicroscopy* **2023**, *246*, 113666.

[110] S. Chen, M. Peng, Y. Li, B.-F. Ju, H. Bao, Y.-L. Chen, G. Zhang, *Commun. Eng.* **2024**, *3*, 131.

[111] I. Sokolov, *Phys. Chem. Chem. Phys.* **2024**, *26*, 11263.

[112] J. Carracedo-Cosme, C. Romero-Muñiz, R. Pérez, *Nanomaterials* **2021**, *11*, 1658.

[113] J. Carracedo-Cosme, C. Romero-Muñiz, P. Pou, R. Pérez, *ACS Appl. Mater. Interfaces* **2023**, *15*, 22692.

[114] S. Farley, J. E. A. Hodgkinson, O. M. Gordon, J. Turner, A. Soltoggio, P. J. Moriarty, E. Hunsicker, *Mach. Learn. Sci. Technol.* **2020**, *2*, 015015.

[115] M. Botifoll, I. Pinto-Huguet, J. Arbiol, *Nanoscale Horizons* **2022**, *7*, 1427.

[116] S. Martí-Sánchez, M. Botifoll, E. Oksenberg, C. Koch, C. Borja, M. C. Spadaro, V. Di Giulio, Q. Ramasse, F. J. García de Abajo, E. Joselevich, J. Arbiol, *Nat. Commun.* **2022**, *13*, 4089.

[117] C.-H. Lee, A. Khan, D. Luo, T. P. Santos, C. Shi, B. E. Janicek, S. Kang, W. Zhu, N. A. Sobh, A. Schleife, B. K. Clark, P. Y. Huang, *Nano Lett.* **2020**, *20*, 3369.

[118] K. Aso, K. Shigematsu, T. Yamamoto, S. Matsumura, *Microscopy* **2016**, *65*, 391.

[119] A. Krull, T.-O. Buchholz, F. Jug, in *2019 IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, IEEE, **2019**, pp. 2124–2132.

[120] J. M. Ede, R. Beanland, *Ultramicroscopy* **2019**, *202*, 18.

[121] J. P. Buban, S.-Y. Choi, *Microsc. Microanal.* **2017**, *23*, 130.

[122] F. Wang, T. R. Henninen, D. Keller, R. Erni, *Appl. Microsc.* **2020**, *50*, 23.

[123] A. Khan, C.-H. Lee, P. Y. Huang, B. K. Clark, *npj Comput. Mater.* **2023**, *9*, 85.

[124] T. M. Quan, D. G. C. Hildebrand, K. Lee, L. A. Thomas, A. T. Kuan, W.-C. A. Lee, W.-K. Jeong, in *2019 IEEE/CVF Int. Conf. Comput. Vis. Work.*, IEEE, **2019**, pp. 3804–3813.

[125] T. Friedrich, C.-P. Yu, J. Verbeeck, S. Van Aert, *Microsc. Microanal.* **2023**, *29*, 395.

[126] S. M. P. Valleti, S. V. Kalinin, C. T. Nelson, J. J. P. Peters, W. Dong, R. Beanland, X. Zhang, I. Takeuchi, M. Ziatdinov, *AIP Adv.* **2022**, *12*, DOI 10.1063/5.0105406.

[127] M. Ziatdinov, O. Dyck, A. Maksov, X. Li, X. Sang, K. Xiao, R. R. Unocic, R. Vasudevan, S. Jesse, S. V. Kalinin, *ACS Nano* **2017**, *11*, 12742.

[128] E. Bladt, D. M. Pelt, S. Bals, K. J. Batenburg, *Ultramicroscopy* **2015**, *158*, 81.

[129] H. Zheng, X. Lu, K. He, *J. Energy Chem.* **2022**, *68*, 454.

[130] S. Anada, Y. Nomura, T. Hirayama, K. Yamamoto, *Microsc. Microanal.* **2020**, *26*, 429.

[131] X. Li, O. Dyck, S. V. Kalinin, S. Jesse, *Microsc. Microanal.* **2018**, *24*, 623.

[132] X. Li, O. Dyck, S. V. Kalinin, S. Jesse, *Microsc. Microanal.* **2019**, *25*, 1688.

[133] C.-H. Lee, D. A. Muller, *Microsc. Microanal.* **2024**, *30*, DOI 10.1093/mam/ozae044.942.

[134] M. J. Cherukara, T. Zhou, Y. Nashed, P. Enfedaque, A. Hexemer, R. J. Harder, M. V. Holt, *Appl. Phys. Lett.* **2020**, *117*, DOI 10.1063/5.0013065.

[135] C. Ophus, *Microsc. Microanal.* **2019**, *25*, 563.

[136] A. R. Lupini, M. P. Oxley, S. V. Kalinin, *Science (80-.).* **2018**, *362*, 399.

[137] K. Kimoto, J. Kikkawa, K. Harano, O. Cretu, Y. Shibazaki, F. Uesugi, *Sci. Rep.* **2024**, *14*, 2901.

[138] H.-C. Ni, R. Yuan, J. Zhang, J.-M. Zuo, *Microsc. Microanal.* **2024**, *30*, DOI 10.1093/mam/ozae044.907.

[139] S. Kandel, S. Maddali, M. Allain, S. O. Hruszkewycz, C. Jacobsen, Y. S. G. Nashed, *Opt. Express* **2019**, *27*, 18653.

[140] M. Hugenschmidt, D. Jannis, A. A. Kadu, L. Grünewald, S. De Marchi, J. Pérez-Juste, J. Verbeeck, S. Van Aert, S. Bals, *ACS Mater. Lett.* **2024**, *6*, 165.

[141] A. Skorikov, W. Heyvaert, W. Albecht, D. M. Pelt, S. Bals, *Nanoscale* **2021**, *13*, 12242.

[142] J. Lee, M. Lee, Y. Park, C. Ophus, Y. Yang, *Phys. Rev. Appl.* **2023**, *19*, 054062.

[143] W. Alvarado, V. Agrawal, W. S. Li, V. P. Dravid, V. Backman, J. J. de Pablo, A. L. Ferguson, *ACS Cent. Sci.* **2023**, *9*, 1200.

[144] S. V. Kalinin, K. M. Roccapriore, S. H. Cho, D. J. Milliron, R. Vasudevan, M. Ziatdinov, J. A. Hachtel, *Adv. Opt. Mater.* **2021**, *9*, DOI 10.1002/adom.202001808.

[145] E. Prifti, J. P. Buban, A. S. Thind, R. F. Klie, *Small* **2023**, *19*, DOI 10.1002/smll.202205977.

[146] S. V. Kalinin, O. Dyck, S. Jesse, M. Ziatdinov, *Sci. Adv.* **2021**, *7*, DOI 10.1126/sciadv.abd5084.

[147] N. Schnitzer, S. H. Sung, R. Hovden, *Microsc. Microanal.* **2019**, *25*, 160.

[148] J. M. Ede, R. Beanland, *Sci. Rep.* **2020**, *10*, 8332.

[149] E. Rotunno, A. H. Tavabi, P. Rosi, S. Frabboni, P. Tiemeijer, R. E. Dunin-Borkowski, V. Grillo, *Ultramicroscopy* **2021**, *228*, 113338.

[150] C. Zhang, Z. Baraissov, C. Duncan, A. Hanuka, A. Edelen, J. Maxson, D. Muller, *Microsc. Microanal.* **2021**, *27*, 810.

[151] E. Rotunno, A. Tavabi, P. Rosi, S. Frabboni, P. Tiemeijer, R. Dunin-Borkovski, V. Grillo, *Microsc. Microanal.* **2021**, *27*, 822.

[152] K. M. Roccapriore, M. G. Boebinger, O. Dyck, A. Ghosh, R. R. Unocic, S. V. Kalinin, M. Ziatdinov, *ACS Nano* **2022**, *16*, 17116.

[153] A. Mittelberger, C. Kramberger, C. Hofer, C. Mangler, J. C. Meyer, *Microsc. Microanal.* **2017**, *23*, 809.

[154] X. Wu, L. Xiao, Y. Sun, J. Zhang, T. Ma, L. He, *Futur. Gener. Comput. Syst.* **2022**, *135*, 364.

[155] S. V Kalinin, Y. Liu, A. Biswas, G. Duscher, U. Pratiush, K. Roccapriore, M. Ziatdinov, R. Vasudevan, *Micros. Today* **2024**, *32*, 35.

[156] A. Gupta, A. Sabirsh, C. Wahlby, I.-M. Sintorn, *IEEE J. Biomed. Heal. Informatics* **2022**, *26*, 4079.

[157] C. Radclyffe, M. Ribeiro, R. H. Wortham, *Front. Artif. Intell.* **2023**, *6*, DOI 10.3389/frai.2023.1020592.

[158] J. Abramson, J. Adler, J. Dunger, R. Evans, T. Green, A. Pritzel, O. Ronneberger, L. Willmore, A. J. Ballard, J. Bambrick, S. W. Bodenstein, D. A. Evans, C.-C. Hung, M. O'Neill, D. Reiman, K. Tunyasuvunakool, Z. Wu, A. Žemgulytė, E. Arvaniti, C. Beattie, O. Bertolli, A. Bridgland, A. Cherepanov, M. Congreve, A. I. Cowen-Rivers, A. Cowie, M. Figurnov, F. B. Fuchs, H. Gladman, R. Jain, Y. A. Khan, C. M. R. Low, K. Perlin, A. Potapenko, P. Savy, S. Singh, A. Stecula, A. Thillaisundaram, C. Tong, S. Yakneen, E. D. Zhong, M. Zielinski, A. Žídek, V. Bapst, P. Kohli, M. Jaderberg, D. Hassabis, J. M. Jumper, *Nature* **2024**, *630*, 493.

[159] E. Erickson, J. E. Gado, L. Avilán, F. Bratti, R. K. Brizendine, P. A. Cox, R. Gill, R. Graham, D.-J. Kim, G. König, W. E. Michener, S. Poudel, K. J. Ramirez, T. J. Shakespeare, M. Zahn, E. S. Boyd, C. M. Payne, J. L. DuBois, A. R. Pickford, G. T. Beckham, J. E. McGeehan, *Nat. Commun.* **2022**, *13*, 7850.

[160] W. Liu, C. Li, B. Li, L. Zhu, D. Ming, L. Jiang, *J. Hazard. Mater.* **2024**, *480*, 136389.

[161] A. Merchant, S. Batzner, S. S. Schoenholz, M. Aykol, G. Cheon, E. D. Cubuk, *Nature* **2023**, *624*, 80.

[162] R. Tran, J. Lan, M. Shuaibi, B. M. Wood, S. Goyal, A. Das, J. Heras-Domingo, A. Kolluru, A. Rizvi, N. Shoghi, A. Sriram, F. Therrien, J. Abed, O. Voznyy, E. H. Sargent, Z. Ulissi, C. L. Zitnick, *ACS Catal.* **2023**, *13*, 3066.

[163] L. Chanussot, A. Das, S. Goyal, T. Lavril, M. Shuaibi, M. Riviere, K. Tran, J. Heras-Domingo, C. Ho, W. Hu, A. Palizhati, A. Sriram, B. Wood, J. Yoon, D. Parikh, C. L. Zitnick, Z. Ulissi, *ACS Catal.* **2021**, *11*, 6059.

[164] "Open Catalyst project," can be found under https://opencatalystproject.org/, **n.d.**

[165] "FairChem," can be found under https://fair-chem.github.io/, **n.d.**

[166] A. Sriram, S. Choi, X. Yu, L. M. Brabson, A. Das, Z. Ulissi, M. Uyttendaele, A. J. Medford, D. S. Sholl, *ACS Cent. Sci.* **2024**, *10*, 923.

[167] L. M. Antunes, K. T. Butler, R. Grau-Crespo, *Nat. Commun.* **2024**, *15*, 10570.

[168] A. M. Bran, S. Cox, O. Schilter, C. Baldassari, A. D. White, P. Schwaller, *Nat. Mach. Intell.* **2024**, *6*, 525.

[169] J. Riebesell, R. E. A. Goodall, P. Benner, Y. Chiang, B. Deng, G. Ceder, M. Asta, A. A. Lee, A. Jain, K. A. Persson, **2023**.

[170] K. Choudhary, K. F. Garrity, A. C. E. Reid, B. DeCost, A. J. Biacchi, A. R. Hight Walker, Z. Trautt, J. Hattrick-Simpers, A. G. Kusne, A. Centrone, A. Davydov, J. Jiang, R. Pachter, G. Cheon, E. Reed, A. Agrawal, X. Qian, V. Sharma, H. Zhuang, S. V. Kalinin, B. G. Sumpter, G. Pilania, P. Acar, S. Mandal, K. Haule, D. Vanderbilt, K. Rabe, F. Tavazza, *npj Comput. Mater.* **2020**, *6*, 173.

[171] Y. Liu, X. Zhang, *Chem. Soc. Rev.* **2011**, *40*, 2494.

[172] N. I. Zheludev, Y. S. Kivshar, *Nat. Mater.* **2012**, *11*, 917.

[173] Y. Zheng, Z. Wu, Eds., *Intelligent Nanotechnology*, Elsevier, **2023**.

[174] G. Cerniauskas, H. Sadia, P. Alam, *Oxford Open Mater. Sci.* **2024**, *4*, DOI 10.1093/oxfmat/itae001.

[175] S. A. Cummer, J. Christensen, A. Alù, *Nat. Rev. Mater.* **2016**, *1*, 16001.

[176] S. H. Lee, C. M. Park, Y. M. Seo, Z. G. Wang, C. K. Kim, *Phys. Lett. A* **2009**, *373*, 4464.

[177] V. M. García-Chocano, R. Graciá-Salgado, D. Torrent, F. Cervera, J. Sánchez-Dehesa, *Phys. Rev. B* **2012**, *85*, 184102.

[178] J. Wang, F. Allein, N. Boechler, J. Friend, O. Vazquez-Mena, *Phys. Rev. Appl.* **2021**, *15*, 024025.

[179] F. Ma, Z. Huang, C. Liu, J. H. Wu, *J. Appl. Phys.* **2022**, *131*, DOI 10.1063/5.0074503.

[180] B. Gibson, T. Nguyen, S. Sinaie, D. Heath, T. Ngo, *Build. Environ.* **2022**, *224*, 109531.

[181] M. Nakayama, *Polym. J.* **2024**, *56*, 71.

[182] M. Ye, L. Gao, H. Li, *Mater. Des.* **2020**, *192*, 108751.

[183] X. Tan, S. Chen, B. Wang, J. Tang, L. Wang, S. Zhu, K. Yao, P. Xu, *Extrem. Mech. Lett.* **2020**, *41*, 100990.

[184] X. Tan, Y. Li, L. Wang, K. Yao, Q. Ji, B. Wang, V. Laude, M. Kadic, *Adv. Intell. Syst.* **2023**, *5*, DOI 10.1002/aisy.202200400.

[185] J. Noronha, J. Dash, J. Rogers, M. Leary, M. Brandt, M. Qian, *Adv. Mater.* **2024**, DOI 10.1002/adma.202308715.

[186] S. Guo, R. Gao, X. Tian, S. Liu, *Eng. Struct.* **2023**, *280*, 115687.

[187] A. Bekele, M. A. Wadee, A. T. M. Phillips, *R. Soc. Open Sci.* **2023**, *10*, DOI 10.1098/rsos.230762.

[188] D. M. J. Dykstra, C. Lenting, A. Masurier, C. Coulais, *Adv. Mater.* **2023**, *35*, DOI 10.1002/adma.202301747.

[189] M. Balan P, J. Mertens A, M. V. A. R. Bahubalendruni, *Mater. Today Commun.* **2023**, *34*, 105285.

[190] M. Kadic, T. Bückmann, R. Schittny, M. Wegener, *Reports Prog. Phys.* **2013**, *76*, 126501.

[191] T. Han, X. Bai, D. Gao, J. Thong, B. Li, C.-W. Qiu, *Phys. Rev. Lett.* **2014**, *112*, 54302.

[192] J. Maire, R. Anufriev, R. Yanagisawa, A. Ramiere, S. Volz, M. Nomura, *Sci. Adv.* **2017**, *3*, e1700027.

[193] H. He, W. Peng, H. Le Ferrand, *Adv. Mater.* **2024**, *36*, DOI 10.1002/adma.202307071.

[194] Y. Li, W. Li, T. Han, X. Zheng, J. Li, B. Li, S. Fan, C.-W. Qiu, *Nat. Rev. Mater.* **2021**, *6*, 488.

[195] R. Hu, W. Xi, Y. Liu, K. Tang, J. Song, X. Luo, J. Wu, C.-W. Qiu, *Mater. Today* **2021**, *45*, 120.

[196] Y. Zhai, Y. Ma, S. N. David, D. Zhao, R. Lou, G. Tan, R. Yang, X. Yin, *Science (80-.).* **2017**, *355*, 1062.

[197] J. Yuan, H. Yin, D. Yuan, Y. Yang, S. Xu, *Energy* **2022**, *242*, 122779.

[198] J. Guo, G. Xu, D. Tian, Z. Qu, C. Qiu, *Adv. Mater.* **2022**, *34*, DOI 10.1002/adma.202201093.

[199] H. Wang, V. Prasad Sivan, A. Mitchell, G. Rosengarten, P. Phelan, L. Wang, *Sol. Energy Mater. Sol. Cells* **2015**, *137*, 235.

[200] E. M. Dede, P. Schmalenberg, T. Nomura, M. Ishigaki, *IEEE Trans. Components, Packag. Manuf. Technol.* **2015**, *5*, 1763.

[201] J.-S. Park, S. W. D. Lim, A. Amirzhan, H. Kang, K. Karrfalt, D. Kim, J. Leger, A. Urbas, M. Ossiander, Z. Li, F. Capasso, *ACS Nano* **2024**, *18*, 3187.

[202] J. Park, J. R. Youn, Y. S. Song, *Phys. Rev. Appl.* **2019**, *12*, 061002.

[203] M. Chen, X. Shen, L. Xu, *Droplet* **2023**, *2*, DOI 10.1002/dro2.79.

[204] F. Hu, W. Wang, J. Cheng, Y. Bao, *Sci. Prog.* **2020**, *103*, DOI 10.1177/0036850420946162.

[205] M. Kadic, G. W. Milton, M. van Hecke, M. Wegener, *Nat. Rev. Phys.* **2019**, *1*, 198.

[206] C. L. Holloway, E. F. Kuester, J. A. Gordon, J. O'Hara, J. Booth, D. R. Smith, *IEEE Antennas*

Propag. Mag. **2012**, *54*, 10.

[207] R. Sarma, M. Goldflam, E. Donahue, A. Pribisova, S. Gennaro, J. Wright, I. Brener, J. Briscoe, *Crystals* **2020**, *10*, 1114.

[208] T. Asano, S. Noda, *Opt. Express* **2018**, *26*, 32704.

[209] D. Nath, Ankit, D. R. Neog, S. S. Gautam, *Arch. Comput. Methods Eng.* **2024**, *31*, 2945.

[210] R. E. Meethal, A. Kodakkal, M. Khalil, A. Ghantasala, B. Obst, K.-U. Bletzinger, R. Wüchner, *Adv. Model. Simul. Eng. Sci.* **2023**, *10*, 6.

[211] I. Malkiel, M. Mrejen, A. Nagler, U. Arieli, L. Wolf, H. Suchowski, *Light Sci. Appl.* **2018**, *7*, 60.

[212] W. Ma, F. Cheng, Y. Liu, *ACS Nano* **2018**, *12*, 6326.

[213] D. Liu, Y. Tan, E. Khoram, Z. Yu, *ACS Photonics* **2018**, *5*, 1365.

[214] Z. Liu, D. Zhu, S. P. Rodrigues, K.-T. Lee, W. Cai, *Nano Lett.* **2018**, *18*, 6570.

[215] W. Ma, F. Cheng, Y. Xu, Q. Wen, Y. Liu, *Adv. Mater.* **2019**, *31*, DOI 10.1002/adma.201901111.

[216] R. Lin, Y. Zhai, C. Xiong, X. Li, *Opt. Lett.* **2020**, *45*, 1362.

[217] S. So, J. Rho, *Nanophotonics* **2019**, *8*, 1255.

[218] E. Ashalley, K. Acheampong, L. V. Besteiro, P. Yu, A. Neogi, A. O. Govorov, Z. M. Wang, *Photonics Res.* **2020**, *8*, 1213.

[219] X. Liao, L. Gui, Z. Yu, T. Zhang, K. Xu, *Opt. Mater. Express* **2022**, *12*, 758.

[220] C. Luo, T. Sang, Z. Ge, J. Lu, Y. Wang, *Opt. Express* **2024**, *32*, 13978.

[221] H. Zheng, Q. Liu, Y. Zhou, I. I. Kravchenko, Y. Huo, J. Valentine, *Sci. Adv.* **2022**, *8*, DOI 10.1126/sciadv.abo6410.

[222] H. Zheng, Q. Liu, I. I. Kravchenko, X. Zhang, Y. Huo, J. G. Valentine, *Nat. Nanotechnol.* **2024**, *19*, 471.

[223] R. E. Christiansen, O. Sigmund, *J. Opt. Soc. Am. B* **2021**, *38*, 496.

[224] J. Jiang, D. Sell, S. Hoyer, J. Hickey, J. Yang, J. A. Fan, *ACS Nano* **2019**, *13*, 8872.

[225] Z. A. Kudyshev, A. V. Kildishev, V. M. Shalaev, A. Boltasseva, *Appl. Phys. Rev.* **2020**, *7*, DOI 10.1063/1.5134792.

[226] J. K. Wilt, C. Yang, G. X. Gu, *Adv. Eng. Mater.* **2020**, *22*, DOI 10.1002/adem.201901266.

[227] S. Bonfanti, S. Hiemer, R. Zulkarnain, R. Guerra, M. Zaiser, S. Zapperi, *Nat. Comput. Sci.* **2024**, *4*, 574.

[228] C. S. Ha, D. Yao, Z. Xu, C. Liu, H. Liu, D. Elkins, M. Kile, V. Deshpande, Z. Kong, M. Bauchy, X. Zheng, *Nat. Commun.* **2023**, *14*, 5765.

[229] C.-X. Liu, G.-L. Yu, G.-Y. Zhao, *AIP Adv.* **2019**, *9*, DOI 10.1063/1.5114643.

[230] X. Li, S. Ning, Z. Liu, Z. Yan, C. Luo, Z. Zhuang, *Comput. Methods Appl. Mech. Eng.* **2020**, *361*, 112737.

[231] A. D. Becke, *J. Chem. Phys.* **2014**, *140*, DOI 10.1063/1.4869598.

[232] W. Kohn, L. J. Sham, *Phys. Rev.* **1965**, *140*, A1133.

[233] N. W. Ashcroft, N. D. Mermin, *Solid State Physics*, Cengage Learning India Private Limited, **2011**.

[234] J. Andzelm, E. Wimmer, *J. Chem. Phys.* **1992**, *96*, 1280.

[235] N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, D. Vanderbilt, *Rev. Mod. Phys.* **2012**, *84*, 1419.

[236] J. M. Soler, E. Artacho, J. D. Gale, A. García, J. Junquera, P. Ordejón, D. Sánchez-Portal, *J. Phys. Condens. Matter* **2002**, *14*, 2745.

[237] J. C. Slater, G. F. Koster, *Phys. Rev.* **1954**, *94*, 1498.

[238] K. F. Garrity, K. Choudhary, *Phys. Rev. Mater.* **2023**, *7*, 044603.

[239] M. Rupp, A. Tkatchenko, K.-R. Müller, O. A. von Lilienfeld, *Phys. Rev. Lett.* **2012**, *108*, 058301.

[240] M. Yaghoobi, M. Alaei, *Comput. Mater. Sci.* **2022**, *207*, 111284.

[241] M. J. van Setten, H. F. W. Dekkers, C. Pashartis, A. Chasin, A. Belmonte, R. Delhougne, G. S. Kar, G. Pourtois, *Mater. Adv.* **2022**, *3*, 8413.

[242] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, M. Sun, *AI Open* **2020**, *1*, 57.

[243] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, P. S. Yu, *IEEE Trans. Neural Networks Learn. Syst.* **2021**, *32*, 4.

[244] T. Xie, J. C. Grossman, *Phys. Rev. Lett.* **2018**, *120*, 145301.

[245] K. T. Schütt, P.-J. Kindermans, H. E. Sauceda, S. Chmiela, A. Tkatchenko, K.-R. Müller, in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.*, Curran Associates Inc., Red Hook, NY, USA, **2017**, pp. 992–1002.

[246] C. Chen, W. Ye, Y. Zuo, C. Zheng, S. P. Ong, *Chem. Mater.* **2019**, *31*, 3564.

[247] K. Choudhary, B. DeCost, *npj Comput. Mater.* **2021**, *7*, 185.

[248] K. Choudhary, B. DeCost, *npj Comput. Mater.* **2022**, *8*, 221.

[249] T. S. Cohen, M. Welling, in *Proc. 33rd Int. Conf. Int. Conf. Mach. Learn. - Vol. 48*, JMLR.Org, **2016**, pp. 2990–2999.

[250] H. Yu, Z. Xu, X. Qian, X. Qian, S. Ji, in *Proc. 40th Int. Conf. Mach. Learn.*, JMLR.Org, **2023**.

[251] J. Nigam, M. J. Willatt, M. Ceriotti, *J. Chem. Phys.* **2022**, *156*, DOI 10.1063/5.0072784.

[252] L. Zhang, B. Onat, G. Dusson, A. McSloy, G. Anand, R. J. Maurer, C. Ortner, J. R. Kermode, *npj Comput. Mater.* **2022**, *8*, 158.

[253] Y. Zhong, H. Yu, M. Su, X. Gong, H. Xiang, *npj Comput. Mater.* **2023**, *9*, 182.

[254] Ö. Almarsson, E. B. Vadas, *Cryst. Growth Des.* **2015**, *15*, 5645.

[255] H. Narayanan, F. Dingfelder, I. Condado Morales, B. Patel, K. E. Heding, J. R. Bjelke, T. Egebjerg, A. Butté, M. Sokolov, N. Lorenzen, P. Arosio, *Mol. Pharm.* **2021**, *18*, 3843.

[256] D. R. Serrano, F. C. Luciano, B. J. Anaya, B. Ongoren, A. Kara, G. Molina, B. I. Ramirez, S. A. Sánchez-Guirales, J. A. Simon, G. Tomietto, C. Rapti, H. K. Ruiz, S. Rawat, D. Kumar, A. Lalatsa, *Pharmaceutics* **2024**, *16*, 1328.

[257] G. Hessler, K.-H. Baringhaus, *Molecules* **2018**, *23*, 2520.

[258] P. Schneider, W. P. Walters, A. T. Plowright, N. Sieroka, J. Listgarten, R. A. Goodnow, J. Fisher, J. M. Jansen, J. S. Duca, T. S. Rush, M. Zentgraf, J. E. Hill, E. Krutoholow, M. Kohler, J. Blaney, K. Funatsu, C. Luebkemann, G. Schneider, *Nat. Rev. Drug Discov.* **2020**, *19*, 353.

[259] A. M. Jordan, *ACS Med. Chem. Lett.* **2018**, *9*, 1150.

[260] M. Schauperl, R. A. Denny, *J. Chem. Inf. Model.* **2022**, *62*, 3142.

[261] D. A. Benson, M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, E. W. Sayers, *Nucleic Acids Res.* **2012**, *41*, D36.

[262] S. Chen, L. C. Francioli, J. K. Goodrich, R. L. Collins, M. Kanai, Q. Wang, J. Alföldi, N. A. Watts, C. Vittal, L. D. Gauthier, T. Poterba, M. W. Wilson, Y. Tarasova, W. Phu, R. Grant, M. T. Yohannes, Z. Koenig, Y. Farjoun, E. Banks, S. Donnelly, S. Gabriel, N. Gupta, S. Ferriera, C. Tolonen, S. Novod, L. Bergelson, D. Roazen, V. Ruano-Rubio, M. Covarrubias, C. Llanwarne, N. Petrillo, G. Wade, T. Jeandet, R. Munshi, K. Tibbetts, M. Abreu, C. A. Aguilar Salinas, T. Ahmad, C. M. Albert, D. Ardissino, I. M. Armean, E. G. Atkinson, G. Atzmon, J. Barnard, S. M. Baxter, L. Beaugerie, E. J. Benjamin, D. Benjamin, M. Boehnke, L. L. Bonnycastle, E. P. Bottinger, D. W. Bowden, M. J. Bown, H. Brand, S. Brant, T. Brookings, S. Bryant, S. E. Calvo, H. Campos, J. C. Chambers, J. C. Chan, K. R. Chao, S. Chapman, D. I. Chasman, R. Chisholm, J. Cho, R. Chowdhury, M. K. Chung, W. K. Chung, K. Cibulskis, B. Cohen, K. M. Connolly, A. Correa, B. B. Cummings, D. Dabelea, J. Danesh, D. Darbar, P. Darnowsky, J. Denny, R. Duggirala, J. Dupuis, P. T. Ellinor, R. Elosua, J. Emery, E. England, J. Erdmann, T. Esko, E. Evangelista, D. Fatkin, J. Florez, A. Franke, J. Fu, M. Färkkilä, K. Garimella, J. Gentry, G. Getz, D. C. Glahn, B. Glaser, S. J. Glatt, D. Goldstein, C. Gonzalez, L. Groop, S. Gudmundsson, A. Haessly, C. Haiman, I. Hall, C. L. Hanis, M. Harms, M. Hiltunen, M. M. Holi, C. M. Hultman, C. Jalas, M. Kallela, D. Kaplan, J. Kaprio, S. Kathiresan, E. E. Kenny, B.-J. Kim, Y. J. Kim, D. King, G. Kirov, J. Kooner, S. Koskinen, H. M. Krumholz, S. Kugathasan, S. H. Kwak, M. Laakso, N. Lake, T. Langsford, K. M. Laricchia, T. Lehtimäki, M. Lek, E. Lipscomb, R. J. F. Loos, W. Lu, S. A. Lubitz, T. T. Luna, R. C. W. Ma, G. M. Marcus, J. Marrugat, K. M. Mattila, S. McCarroll, M. I. McCarthy, J. L. McCauley, D. McGovern, R. McPherson, J. B. Meigs, O. Melander, A. Metspalu, D. Meyers, E. V. Minikel, B. D. Mitchell, V. K. Mootha, A. Naheed, S. Nazarian, P. M. Nilsson, M. C. O'Donovan, Y. Okada, D. Ongur, L. Orozco, M. J. Owen, C. Palmer, N. D. Palmer, A. Palotie, K. S. Park, C. Pato, A. E. Pulver, D. Rader, N. Rahman, A. Reiner, A. M. Remes, D. Rhodes, S. Rich, J. D. Rioux, S. Ripatti, D. M. Roden, J. I. Rotter, N. Sahakian, D. Saleheen, V. Salomaa, A. Saltzman, N. J. Samani, K. E. Samocha, A. Sanchis-Juan, J. Scharf, M. Schleicher, H. Schunkert, S. Schönherr, E. G. Seaby, S. H. Shah, M. Shand, T. Sharpe, M. B. Shoemaker, T. Shyong, E. K. Silverman, M. Singer-Berk, P. Sklar, J. T. Smith, J. G. Smith, H. Soininen, H. Sokol, R. G. Son, J. Soto, T. Spector, C. Stevens, N. O. Stitziel, P. F. Sullivan, J. Suvisaari, E. S. Tai, K. D. Taylor, Y. Y. Teo, M. Tsuang, T. Tuomi, D. Turner, T. Tusie-Luna, E. Vartiainen, M. Vawter, L. Wang, A. Wang, J. S. Ware, H. Watkins, R. K. Weersma, B. Weisburd, M. Wessman, N. Whiffin, J. G. Wilson, R. J. Xavier, A. O'Donnell-Luria, M. Solomonson, C. Seed, A. R. Martin, M. E. Talkowski, H. L. Rehm, M. J. Daly, G. Tiao, B. M. Neale, D. G. MacArthur, K. J. Karczewski, *Nature* **2024**, *625*, 92.

[263] S. Chang, J.-Y. Chen, Y.-J. Chuang, B.-S. Chen, *Int. J. Mol. Sci.* **2020**, *22*, 166.

[264] L. Chi, J. Ma, Y. Wan, Y. Deng, Y. Wu, X. Cen, X. Zhou, X. Zhao, Y. Wang, Z. Ji, *bioRxiv: 2023.12.10.571021* **2023**, DOI 10.1101/2023.12.10.571021.

[265] X. Zeng, S. Zhu, W. Lu, Z. Liu, J. Huang, Y. Zhou, J. Fang, Y. Huang, H. Guo, L. Li, B. D. Trapp, R. Nussinov, C. Eng, J. Loscalzo, F. Cheng, *Chem. Sci.* **2020**, *11*, 1775.

[266] D. Desai, S. V Kantliwala, J. Vybhavi, R. Ravi, H. Patel, J. Patel, *Cureus* **2024**, DOI 10.7759/cureus.63646.

[267] J. Wee, G.-W. Wei, *J. Chem. Inf. Model.* **2024**, *64*, 6676.

[268] Anusha, Z. Zhang, J. Li, H. Zuo, C. Mao, *J. Am. Chem. Soc.* **2024**, *146*, 25422.

[269] S. Tsuji, T. Hase, A. Yachie-Kinoshita, T. Nishino, S. Ghosh, M. Kikuchi, K. Shimokawa, H. Aburatani, H. Kitano, H. Tanaka, *Alzheimers. Res. Ther.* **2021**, *13*, 92.

[270] H. Kurata, S. Tsukiyama, *PLoS One* **2022**, *17*, e0276609.

[271] D. Jones, H. Kim, X. Zhang, A. Zemla, G. Stevenson, W. F. D. Bennett, D. Kirshner, S. E. Wong, F. C. Lightstone, J. E. Allen, *J. Chem. Inf. Model.* **2021**, *61*, 1583.

[272] S. Limbu, S. Dakshanamurthy, *Int. J. Mol. Sci.* **2022**, *23*, 13912.

[273] A. Ahmed, B. Mam, R. Sowdhamini, *Bioinform. Biol. Insights* **2021**, *15*, 117793222110303.

[274] B. Ma, K. Terayama, S. Matsumoto, Y. Isaka, Y. Sasakura, H. Iwata, M. Araki, Y. Okuno, *J. Chem. Inf. Model.* **2021**, *61*, 3304.

[275] P. Bannigan, M. Aldeghi, Z. Bao, F. Häse, A. Aspuru-Guzik, C. Allen, *Adv. Drug Deliv. Rev.* **2021**, *175*, 113806.

[276] A. T, R. Narayan, P. A. Shenoy, U. Y. Nayak, *J. Mol. Liq.* **2022**, *368*, 120596.

[277] S. A. Damiati, L. G. Martini, N. W. Smith, M. J. Lawrence, D. J. Barlow, *Int. J. Pharm.* **2017**, *530*, 99.

[278] D. Leonardi, C. J. Salomón, M. C. Lamas, A. C. Olivieri, *Int. J. Pharm.* **2009**, *367*, 140.

[279] Y. Li, M. R. Abbaspour, P. V. Grootendorst, A. M. Rauth, X. Y. Wu, *Eur. J. Pharm. Biopharm.* **2015**, *94*, 170.

[280] A. Ghosh, L. Louis, K. K. Arora, B. C. Hancock, J. F. Krzyzaniak, P. Meenan, S. Nakhmanson, G. P. F. Wood, *CrystEngComm* **2019**, *21*, 1215.

[281] Y. Ma, Z. Gao, P. Shi, M. Chen, S. Wu, C. Yang, J. Wang, J. Cheng, J. Gong, *Front. Chem. Sci. Eng.* **2022**, *16*, 523.

[282] Q. Zhao, Z. Ye, Y. Su, D. Ouyang, *Acta Pharm. Sin. B* **2019**, *9*, 1241.

[283] B. Hoseini, M. R. Jaafari, A. Golabpour, A. A. Momtazi-Borojeni, M. Karimi, S. Eslami, *Sci. Rep.* **2023**, *13*, 18012.

[284] K. P. Das, C. J, *Front. Med. Technol.* **2023**, *4*, DOI 10.3389/fmedt.2022.1067144.

[285] M. Kuhn, I. Letunic, L. J. Jensen, P. Bork, *Nucleic Acids Res.* **2016**, *44*, D1075.

[286] P. Banerjee, E. Kemmler, M. Dunkel, R. Preissner, *Nucleic Acids Res.* **2024**, *52*, W513.

[287] A. Mayr, G. Klambauer, T. Unterthiner, S. Hochreiter, *Front. Environ. Sci.* **2016**, *3*, DOI 10.3389/fenvs.2015.00080.

[288] A.-K. Sohlenius-Sternbeck, Y. Terelius, *Drug Metab. Dispos.* **2022**, *50*, 95.

[289] F. Sanz, F. Pognan, T. Steger-Hartmann, C. Díaz, M. Cases, M. Pastor, P. Marc, J. Wichard, K. Briggs, D. K. Watson, T. Kleinöder, C. Yang, A. Amberg, M. Beaumont, A. J. Brookes, S. Brunak, M. T. D. Cronin, G. F. Ecker, S. Escher, N. Greene, A. Guzmán, A. Hersey, P. Jacques, L. Lammens, J. Mestres, W. Muster, H. Northeved, M. Pinches, J. Saiz, N. Sajot, A. Valencia, J. van der Lei, N. P. E. Vermeulen, E. Vock, G. Wolber, I. Zamora, *Nat. Rev. Drug Discov.* **2017**, *16*, 811.

[290] M. Cases, K. Briggs, T. Steger-Hartmann, F. Pognan, P. Marc, T. Kleinöder, C. Schwab, M. Pastor, J. Wichard, F. Sanz, *Int. J. Mol. Sci.* **2014**, *15*, 21136.

[291] M. Di Stefano, S. Galati, L. Piazza, C. Granchi, S. Mancini, F. Fratini, M. Macchia, G. Poli, T. Tuccinardi, *J. Chem. Inf. Model.* **2024**, *64*, 2275.

[292] D. Naga, S. Dimitrakopoulou, S. Roberts, E. Husar, S. Mohr, H. Booler, E. Musvasva, *Sci. Rep.* **2023**, *13*, 14865.

[293] Y. Cha, T. Erez, I. J. Reynolds, D. Kumar, J. Ross, G. Koytiger, R. Kusko, B. Zeskind, S. Risso, E. Kagan, S. Papapetropoulos, I. Grossman, D. Laifenfeld, *Br. J. Pharmacol.* **2018**, *175*, 168.

[294] R. Y. Booq, E. A. Tawfik, H. A. Alfassam, A. J. Alfahad, E. J. Alyamani, *Antibiotics* **2021**, *10*, 1480.

[295] S. Higashihira, S. J. Simpson, C. D. Collier, R. M. Natoli, M. Kittaka, E. M. Greenfield, *Clin. Orthop. Relat. Res.* **2022**, *480*, 1476.

[296] Z. Hussain, S. Pengfei, L. Yimin, L. Shasha, L. Zehao, Y. Yifan, L. Linhui, Z. Linying, W. Yong, *Pathog. Dis.* **2022**, *80*, DOI 10.1093/femspd/ftac037.

[297] P. Richardson, I. Griffin, C. Tucker, D. Smith, O. Oechsle, A. Phelan, M. Rawling, E. Savory, J. Stebbing, *Lancet* **2020**, *395*, e30.

[298] J. Stebbing, V. Krishnan, S. de Bono, S. Ottaviani, G. Casalini, P. J. Richardson, V. Monteil, V. M. Lauschke, A. Mirazimi, S. Youhanna, Y. Tan, F. Baldanti, A. Sarasini, J. A. R. Terres, B. J. Nickoloff, R. E. Higgs, G. Rocha, N. L. Byers, D. E. Schlichting, A. Nirula, A. Cardoso, M. Corbellino, *EMBO Mol. Med.* **2020**, *12*, DOI 10.15252/emmm.202012697.

[299] A. Zhavoronkov, Y. A. Ivanenkov, A. Aliper, M. S. Veselov, V. A. Aladinskiy, A. V. Aladinskaya, V. A. Terentiev, D. A. Polykovskiy, M. D. Kuznetsov, A. Asadulaev, Y. Volkov, A. Zholus, R. R. Shayakhmetov, A. Zhebrak, L. I. Minaeva, B. A. Zagribelnyy, L. H. Lee, R. Soll, D. Madge, L. Xing, T. Guo, A. Aspuru-Guzik, *Nat. Biotechnol.* **2019**, *37*, 1038.

[300] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, Q. He, *Proc. IEEE* **2021**, *109*, 43.

[301] V. Gupta, K. Choudhary, F. Tavazza, C. Campbell, W. Liao, A. Choudhary, A. Agrawal, *Nat. Commun.* **2021**, *12*, 6595.

[302] R. Meyes, M. Lu, C. W. de Puiseau, T. Meisen, **2019**.

[303] M. D. Zeiler, R. Fergus, **2014**, pp. 818–833.

[304] M. T. Ribeiro, S. Singh, C. Guestrin, in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, ACM, New York, NY, USA, **2016**, pp. 1135–1144.

[305] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, *Int. J. Comput. Vis.* **2020**, *128*, 336.

[306] J. Dastin, S. Nellis, *Reuters* **2023**.

[307] D. Patel, A. Ahmad, "The Inference Cost Of Search Disruption – Large Language Model Cost Analysis," can be found under https://www.semianalysis.com/p/the-inference-cost-of-search-disruption, **2023**.

[308] A. de Vries, *Joule* **2023**, *7*, 2191.

[309] W. Maass, *Neural Networks* **1997**, *10*, 1659.

[310] Y. Kim, Y. Li, H. Park, Y. Venkatesha, P. Panda, **2022**, pp. 36–56.

[311] R. Ciancio, R. E. Dunin-Borkowski, E. Snoeck, M. Kociak, R. Holmestad, J. Verbeeck, A. I. Kirkland, G. Kothleitner, J. Arbiol, *Microsc. Microanal.* **2022**, *28*, 2900.