Scientific discovery in the age of artificial intelligence

https://doi.org/10.1038/s41586-023-06221-2

Received: 30 March 2022

Accepted: 16 May 2023

Published online: 2 August 2023

Check for updates

Hanchen Wang[1,2,37,38,39], Tianfan Fu[3,39], Yuanqi Du[4,39], Wenhao Gao[5], Kexin Huang[6], Ziming Liu[7], Payal Chandak[8], Shengchao Liu[9,10], Peter Van Katwyk[11,12], Andreea Deac[9,10], Anima Anandkumar[2,13], Karianne Bergen[11,12], Carla P. Gomes[4], Shirley Ho[14,15,16,17], Pushmeet Kohli[18], Joan Lasenby[1], Jure Leskovec[6], Tie-Yan Liu[19], Arjun Manrai[20], Debora Marks[21,22], Bharath Ramsundar[23], Le Song[24,25], Jimeng Sun[26], Jian Tang[9,27,28], Petar Veličković[17,29], Max Welling[30,31], Linfeng Zhang[32,33], Connor W. Coley[5,34], Yoshua Bengio[9,10] & Marinka Zitnik[20,22,35,36] ✉

Artificial intelligence (AI) is being increasingly integrated into scientific discovery to augment and accelerate research, helping scientists to generate hypotheses, design experiments, collect and interpret large datasets, and gain insights that might not have been possible using traditional scientific methods alone. Here we examine breakthroughs over the past decade that include self-supervised learning, which allows models to be trained on vast amounts of unlabelled data, and geometric deep learning, which leverages knowledge about the structure of scientific data to enhance model accuracy and efficiency. Generative AI methods can create designs, such as small-molecule drugs and proteins, by analysing diverse data modalities, including images and sequences. We discuss how these methods can help scientists throughout the scientific process and the central issues that remain despite such advances. Both developers and users of AI toolsneed a better understanding of when such approaches need improvement, and challenges posed by poor data quality and stewardship remain. These issues cut across scientific disciplines and require developing foundational algorithmic approaches that can contribute to scientific understanding or acquire it autonomously, making them critical areas of focus for AI innovation.

The foundation for forming scientific insights and theories is laid by how data are collected, transformed and understood. The rise of deep learning in the early 2010s has significantly expanded the scope and ambition of these scientific discovery processes[1]. Artificial intelligence (AI) is increasingly used across scientific disciplines to integrate massive datasets, refine measurements, guide experimentation, explore the space of theories compatible with the data, and provide actionable and reliable models integrated with scientific workflows for autonomous discovery.

Data collection and analysis are fundamental to scientific understanding and discovery, two of the central aims in science[2], and quantitative methods and emerging technologies, from physical instruments such as microscopes to research techniques such as bootstrapping, have long been used to reach these aims[3]. The introduction of digitization in the 1950s paved the way for the general use of computing in scientific research. The rise of data science since the 2010s has enabled AI to provide valuable guidance by identifying scientifically relevant patterns from large datasets.

Although scientific practices and procedures vary across stages of scientific research, the development of AI algorithms cuts across traditionally isolated disciplines (Fig. 1). Such algorithms can enhance the design and execution of scientific studies. They are becoming

[1]Department of Engineering, University of Cambridge, Cambridge, UK. [2]Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA, USA. [3]Department of Computational Science and Engineering, Georgia Institute of Technology, Atlanta, GA, USA. [4]Department of Computer Science, Cornell University, Ithaca, NY, USA. [5]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [6]Department of Computer Science, Stanford University, Stanford, CA, USA. [7]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA, USA. [8]Harvard-MIT Program in Health Sciences and Technology, Cambridge, MA, USA. [9]Mila – Quebec AI Institute, Montreal, Quebec, Canada. [10]Université de Montréal, Montreal, Quebec, Canada. [11]Department of Earth, Environmental and Planetary Sciences, Brown University, Providence, RI, USA. [12]Data Science Institute, Brown University, Providence, RI, USA. [13]NVIDIA, Santa Clara, CA, USA. [14]Center for Computational Astrophysics, Flatiron Institute, New York, NY, USA. [15]Department of Astrophysical Sciences, Princeton University, Princeton, NJ, USA. [16]Department of Physics, Carnegie Mellon University, Pittsburgh, PA, USA. [17]Department of Physics and Center for Data Science, New York University, New York, NY, USA. [18]Google DeepMind, London, UK. [19]Microsoft Research, Beijing, China. [20]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. [21]Department of Systems Biology, Harvard Medical School, Boston, MA, USA. [22]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [23]Deep Forest Sciences, Palo Alto, CA, USA. [24]BioMap, Beijing, China. [25]Mohamed bin Zayed University of Artificial Intelligence, Abu Dhabi, United Arab Emirates. [26]University of Illinois at Urbana-Champaign, Champaign, IL, USA. [27]HEC Montréal, Montreal, Quebec, Canada. [28]CIFAR AI Chair, Toronto, Ontario, Canada. [29]Department of Computer Science and Technology, University of Cambridge, Cambridge, UK. [30]University of Amsterdam, Amsterdam, Netherlands. [31]Microsoft Research Amsterdam, Amsterdam, Netherlands. [32]DP Technology, Beijing, China. [33]AI for Science Institute, Beijing, China. [34]Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, USA. [35]Harvard Data Science Initiative, Cambridge, MA, USA. [36]Kempner Institute for the Study of Natural and Artificial Intelligence, Harvard University, Cambridge, MA, USA. [37]Present address: Department of Research and Early Development, Genentech Inc, South San Francisco, CA, USA. [38]Present address: Department of Computer Science, Stanford University, Stanford, CA, USA. [39]These authors contributed equally: Hanchen Wang, Tianfan Fu, Yuanqi Du. ✉e-mail: marinka@hms.harvard.edu

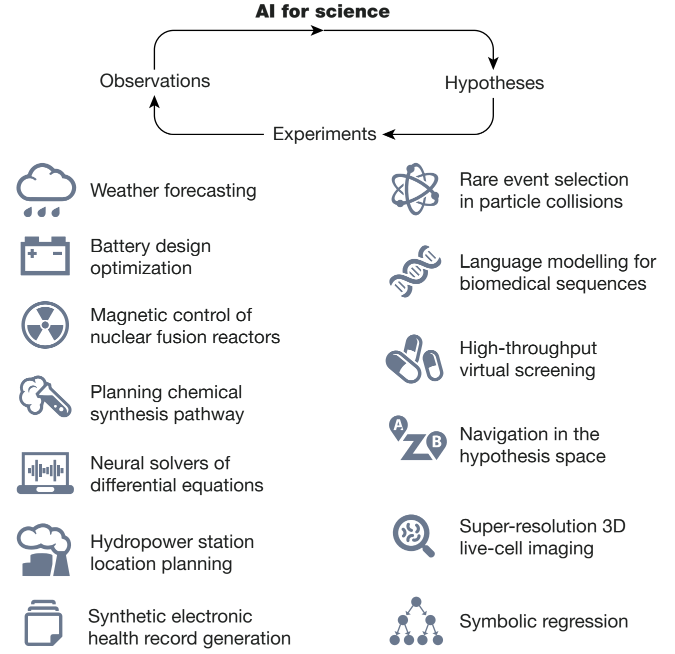
AI for science
Observations
Hypotheses
Experiments
Weather forecasting
Battery design optimization
Magnetic control of nuclear fusion reactors
Planning chemical synthesis pathway
Neural solvers of differential equations
Hydropower station location planning
Synthetic electronic health record generation
Rare event selection in particle collisions
Language modelling for biomedical sequences
High-throughput virtual screening
Navigation in the hypothesis space
Super-resolution 3D live-cell imaging
Symbolic regression

Fig. 1 | Science in the age of artificial intelligence. Scientific discovery is a multifaceted process that involves several interconnected stages, including hypothesis formation, experimental design, data collection and analysis. AI is poised to reshape scientific discovery by augmenting and accelerating research at each stage of this process. The principles and illustrative studies shown here highlight the contributions to enhance scientific understanding and discovery.

indispensable tools for researchers by optimizing parameters and functions[4], automating procedures to collect, visualize, and process data[5], exploring vast spaces of candidate hypotheses to form theories[6], and generating hypotheses and estimating their uncertainty to suggest relevant experiments[7].

The power of AI methods has vastly increased since the early 2010s because of the availability of large datasets, aided by fast and massively parallel computing and storage hardware (graphics processing units and supercomputers) and coupled with new algorithms. The latter includes deep representation learning (Box 1), particularly multilayered neural networks capable of identifying essential, compact features that can simultaneously solve many tasks that underlie a scientific problem. Of these, geometric deep learning (Box 1) has proved to be helpful in integrating scientific knowledge, presented as compact mathematical statements of physical relationships, prior distributions, constraints and other complex descriptors, such as the geometry of atoms in molecules. Self-supervised learning (Box 1) has enabled neural networks trained on labelled or unlabelled data to transfer learned representations to a different domain with few labelled examples, for example, by pre-training large foundation models[8] and adapting them to solve diverse tasks across different domains. In addition, generative models (Box 1) can estimate the underlying data distribution of a complex system and support new designs. Distinct from other uses of AI, reinforcement-learning methods (Box 1) find optimal strategies for an environment by exploring many possible scenarios and assigning rewards to different actions based on metrics such as the information gain expected from a considered experiment.

In AI-driven scientific discovery, scientific knowledge can be incorporated into AI models using appropriate inductive biases (Box 1), which are assumptions representing structure, symmetry, constraints and prior knowledge as compact mathematical statements. However, applying these laws can lead to equations that are too complex for humans to solve, even with traditional numerical methods[9]. An emerging approach is incorporating scientific knowledge into AI models by including information about fundamental equations, such as the laws of physics or principles of molecular structure and binding in protein folding. Such inductive biases can enhance AI models by reducing the number of training examples needed to achieve the same level of accuracy[10] and scaling analyses to a vast space of unexplored scientific hypotheses[11].

Using AI for scientific innovation and discovery presents unique challenges compared with other areas of human endeavour where AI is utilized. One of the biggest challenges is the vastness of hypothesis spaces in scientific problems, making systematic exploration infeasible. For instance, in biochemistry, an estimated 10^{60} drug-like molecules exist to explore[12]. AI systems have the potential to revolutionize scientific workflows by accelerating processes and providing predictions with near-experimental accuracy. However, there are challenges to obtaining reliably annotated datasets for AI models, which can involve time-consuming and resource-intensive experimentation and simulations[13]. Despite these challenges, AI systems can enable efficient, intelligent and highly autonomous experimental design and data collection, where AI systems can operate under human supervision to assess, evaluate and act on results. Such capabilities have facilitated the development of artificially intelligent agents that continuously interact in dynamic environments and can, for example, make real-time decisions to navigate stratospheric balloons[14]. AI systems can play a valuable role in interpreting scientific datasets and extracting relationships and knowledge from scientific literature in a generalized manner. Recent findings demonstrate the potential for unsupervised language AI models to capture complex scientific concepts[15], such as the periodic table, and predict applications of functional materials years before their discovery, suggesting that latent knowledge regarding future discoveries may be embedded in past publications.

Recent advances, including the successful unraveling of the 50-year-old protein-folding problem[10] and AI-driven simulations of molecular systems with millions of particles[16], demonstrate the potential of AI to address challenging scientific problems. However, the remarkable promise of discovery is accompanied by significant challenges for the emerging field of 'AI for Science' (AI4Science). As with any new technology, the success of AI4Science depends on our ability to integrate it into routine practices and understand its potential and limitations. Barriers to the widespread adoption of AI in scientific discovery include internal and external factors specific to each stage of the discovery process and concerns regarding the utility of methods, theory, software and hardware, as well as potential misuse. We explore the developments and address critical questions in AI4Science, including the conduct of science, traditional scepticism and implementation challenges.

AI-aided data collection and curation for scientific research

The ever-increasing scale and complexity of datasets collected by experimental platforms have led to a growing dependence on real-time processing and high-performance computing in scientific research to selectively store and analyse data generated at high rates[17].

Data selection

A typical particle collision experiment generates over 100 terabytes of data every second[18]. Such scientific experiments are pushing the limits of existing data transmission and storage technologies. In these physics experiments, over 99.99% of raw instrument data represents background events that must be detected in real time and discarded to manage the data rates[18]. To identify rare events for future scientific enquiry, deep-learning methods[18] replace pre-programmed hardware event triggers with algorithms that search for outlying signals to detect unforeseen or rare phenomena that may otherwise

Box 1

Glossary

Active learning can improve AI models by selecting the most informative training points when data labelling is costly. Bayesian optimization is a sequential strategy used for optimizing expensive black-box functions and often works with active learning to determine the next query to the black-box function.

An autoencoder is a neural architecture that learns a compressed representation of unlabelled data, consisting of an encoder (which maps data to a representation) and a decoder (which reconstructs data from the representation).

Data augmentation is a strategy that enhances model robustness and generalizability by creating new data samples from existing ones. This process can involve substituting tokens in a sequence, altering the visual aspects of images, or changing atomic positions, always to preserve the essential information. This technique not only increases the diversity of the data but also its volume, thereby aiding in the training of models.

Distribution shift is a prevalent issue in the application of AI methods, whereby the underlying data distribution that an algorithm was initially trained on differs from the distribution of the data it encounters during implementation.

End-to-end learning uses differentiable components, such as neural network modules, to directly connect raw inputs to outputs, avoiding the need for handcrafted input features and enabling direct generation of predictions from inputs.

Generative models estimate a probability distribution of the underlying data and can then generate new samples from that distribution. Examples include variational autoencoders, generative adversarial networks, normalizing flows, diffusion models and generative pretrained transformers.

Geometric deep learning is a field of machine learning that deals with geometric data, such as graphs or manifolds. It typically preserves the invariance of geometric data under transformations and can be applied to 3D structures.

Inductive bias refers to a set of assumptions or preferences that guide the decision-making process of AI models, such as translation equivariance in convolutional networks.

An inverse problem is a scientific or mathematical challenge where the goal is to decipher the underlying causes or parameters that resulted in a specific observation or dataset. Instead of a direct, forward prediction from cause to effect, inverse problems operate in the opposite direction, seeking to deduce the original conditions from the resulting observation. These problems are often complicated due to non-uniqueness and instability, where multiple sets of causes can lead to similar outcomes and minor changes in data can drastically alter the solution.

Physics-informed AI refers to techniques that incorporate physical laws into AI models as a form of prior knowledge.

Reinforcement learning involves sequential decision-making and is represented as a Markov decision process comprising an agent, a set of states, a space of actions, an environment (which determines how the state changes with actions) and a reward function. The reinforcement-learning agent is trained to choose optimal actions based on a state that results in the maximum expected cumulative reward.

Representation learning techniques automatically generate representations of data such as images, documents, sequences or graphs. These representations are typically dense, compact vectors, referred to as embeddings or latent vectors, optimized to capture essential features of input data.

Self-supervised learning is a training strategy for learning from unlabelled data. Generative self-supervised learning, for example, involves predicting a part of the raw data based on the rest, whereas contrastive self-supervised learning involves defining positive and negative views of the input and then aligning the positives and separating the negatives. Both approaches aim to enhance the model's ability to learn meaningful features without needing labelled data.

Surrogate models are analytically tractable models to approximate properties of complex systems.

Symmetries. Equivariance, also known as covariance in physics, characterizes the symmetry of functions. An equivariant function transforms the input equivalently under an operation from a particular group. Invariance is another form of symmetry where a function is invariant to a group of transformations if the output remains unchanged when the inputs are transformed.

A transformer is a neural architecture that uses attention for parallel processing of sequential data via a series of steps. At every step, the attention mechanism selects and combines elements from the previous-step sequence, forming a new representation for each position in the sequence in a differentiable and soft manner.

Weakly supervised learning leverages imperfect, partial or noisy forms of supervision, such as biased or imprecise labels, to train AI models.

be missed during compression. The background processes can be modelled generatively using a deep autoencoder[19] (Box 1). The autoencoder[20] returns a higher loss value (anomaly score) for previously unseen signals (rare events) that are out of the background distribution. Unlike supervised anomaly detection, unsupervised anomaly detection does not require annotations and has been widely used in physics[21,22], neuroscience[23], Earth science[24], oceanography[25] and astronomy[26].

Data annotation

Training supervised models requires datasets with annotated labels that provide supervised information to guide model training and estimate a function or a conditional distribution over target variables from inputs. Pseudo-labelling[27] and label propagation[28] are enticing alternatives to laborious data labelling, allowing automatic annotation of massive unlabelled datasets based on only a small set of accurate annotations. In biology, techniques that assign functional and structural labels to newly characterized molecules are vital for downstream training of supervised models owing to the difficulty of experimentally generating labels. For instance, less than 1% of sequenced proteins is annotated with biological functions despite the proliferation of next-generation sequencing[29]. Another strategy for data labelling leverages surrogate models trained on manually labelled data to annotate unlabelled samples and uses these predicted pseudo-labels to supervise downstream predictive models. In contrast, label propagation diffuses labels to unlabelled samples via similarity graphs constructed based on feature embeddings[13,30] (Box 1). In addition to automatic labelling, active learning[31–33] (Box 1) can identify the most informative data points to be labelled by humans or the most informative experiments to be performed. This approach allows models to be trained with fewer expert-provided labels. Another strategy in data annotation is to develop labelling rules that leverage domain knowledge[34,35].

Data generation

Deep-learning performance improves with increased quality, diversity and scale[36] of training datasets[37,38]. A fruitful approach to creating better models is to augment training datasets by generating additional synthetic data points through automatic data augmentation and deep generative models. In addition to manually designing such data augmentations (Box 1), reinforcement-learning methods[39] can discover a policy for automatic data augmentation[40,41] that is flexible and agnostic of downstream models. Deep generative models, including variational autoencoders, generative adversarial networks, normalizing flows and diffusion models, learn the underlying data distribution and can sample training points from the optimized distribution. Generative adversarial networks (Box 1) have proven to be beneficial for scientific images because they synthesize realistic images in many domains ranging from particle collision events[42], pathology slides[43], chest X-rays[44], magnetic resonance contrasts[45], three-dimensional (3D) material microstructure[46], protein functions[47,48] to genetic sequences[49]. An emerging technique in generative modelling is probabilistic programming[50], in which data generation models are expressed as computer programs.

Data refinements

Precision instruments such as ultrahigh-resolution lasers and non-invasive microscopy systems enable direct measurement of physical quantities or indirect measurement by calculating real-world objects, producing highly accurate results. AI techniques have significantly increased measurement resolution, reduced noise and eliminated errors in measuring roundness, resulting in high precision consistent across sites. Examples of AI applications in scientific experiments include visualization regions of spacetime such as black holes[5], capturing a physics particle collision[51], improving the resolution of live-cell images[52] and better detection of cell types across biological contexts[53]. Deep convolutional methods, which utilize algorithmic advances such as spectral deconvolution[54,55], flexible sparsity[52] and generative capability[56], can transform poor spatiotemporally resolved measurements into high-quality, super-resolved and structured images. An important AI task in various scientific disciplines is denoising, which involves differentiating relevant signals from noise and learning to remove noise. Denoising autoencoders[57] can project high-dimensional input data into more compact representations of essential features. These autoencoders minimize the difference between uncorrupted input data points and their reconstruction from the compressed representation of their noise-corrupted version. Other forms of distribution-learning autoencoders, such as variational autoencoders (VAEs; Box 1)[58], are also frequently used. VAEs learn a stochastic representation via latent autoencoding that retains essential data features while ignoring non-essential sources of variation, probably representing random noise. For example, in single-cell genomics, autoencoders optimizing count-based vectors of gene activation across millions of cells[59] are routinely used to improve protein-RNA expression analyses.

Learning meaningful representations of scientific data

Deep learning can extract meaningful representations of scientific data at various levels of abstraction and optimize them to guide research, often through end-to-end learning (Box 1). A high-quality representation should retain as much information about the data as possible while remaining simple and accessible[60]. Scientifically meaningful representations are compact[21], discriminative[61], disentangle underlying factors of variation[62] and encode underlying mechanisms that generalize across numerous tasks[63,64]. Here we introduce three emerging strategies that fulfil these requirements: geometric priors, self-supervised learning and language modelling.

Geometric priors

Integrating geometric priors[65] in learned representations has proved effective as geometry and structure play a central role in scientific domains[66–68]. Symmetry is a widely studied concept in geometry[69]. It can be described in terms of invariance and equivariance (Box 1) to represent the behaviour of a mathematical function, such as a neural feature encoder, under a group of transformations, such as the SE(3) group in rigid body dynamics. Important structural properties, such as the secondary structural content of molecular systems, solvent accessibility, residue compactness and hydrogen-bonding patterns, are invariant to spatial orientations. In the analysis of scientific images, objects do not change when translated in the image, meaning that image segmentation masks are translationally equivariant as they change equivalently when input pixels are translated. Incorporating symmetry into models can benefit using AI with limited labelled datasets, such as 3D RNA and protein structures[70,71], by augmenting training samples, and can improve extrapolative prediction to inputs markedly different than those encountered during model training.

Geometric deep learning

Graph neural networks have emerged as a leading approach for deep learning on datasets with underlying geometric and relational structure[72–76] (Fig. 2a). In a broader sense, geometric deep learning involves discovering relational patterns[65] and equipping neural network models with inductive biases that explicitly make use of localized information encoded in the form of graphs and transformation groups[77–79] through neural message-passing algorithms[80–84]. Depending on the scientific problem, various graph representations were developed to capture complex systems[85–87]. Directional edges can facilitate the physical modelling of glassy systems[88], hypergraphs with edges connecting multiple nodes are used in chromatin structure understanding[89], models trained on multimodal graphs are used to create predictive models in genomics[90], and sparse, irregular and highly relational graphs have been applied to a number of Large Hadron Collider physics tasks, including the reconstruction of particles from detector readouts and the discrimination of physics signals against background processes[91].

Self-supervised learning

Supervised learning may be insufficient when only a few labelled samples are available for model training or when labelling data for a specific task is prohibitively expensive. In such cases, leveraging both labelled and unlabelled data can improve model performance and learning capacity. Self-supervised learning is a technique that enables models to learn the general features of a dataset without relying on explicit labels. Effective self-supervised strategies include predicting occluded regions of images, forecasting past or future frames in a video, and using contrastive learning to teach the model to distinguish between similar and dissimilar data points[92] (Fig. 2b). Self-supervised learning can be a crucial pre-processing step to learn transferable features in large unlabelled datasets[92–95] before fine-tuning models on small labelled datasets to perform downstream tasks. Such pretrained models[96–98] with a broad understanding of a scientific domain are general-purpose predictors that can be adapted for various tasks, thereby improving label efficiency and surpassing purely supervised methods[8].

Language modelling

Masked-language modelling is a popular method for self-supervised learning of both natural language and biological sequences (Fig. 2c). The arrangement of atoms or amino acids (tokens) into structures to produce molecular and biological function is similar to how letters form words and sentences to define the meaning of a document. As both natural language and biological sequence processing continue to evolve, they inform the development of each other. In the training process, the goal is to predict the next token in a sequence, whereas

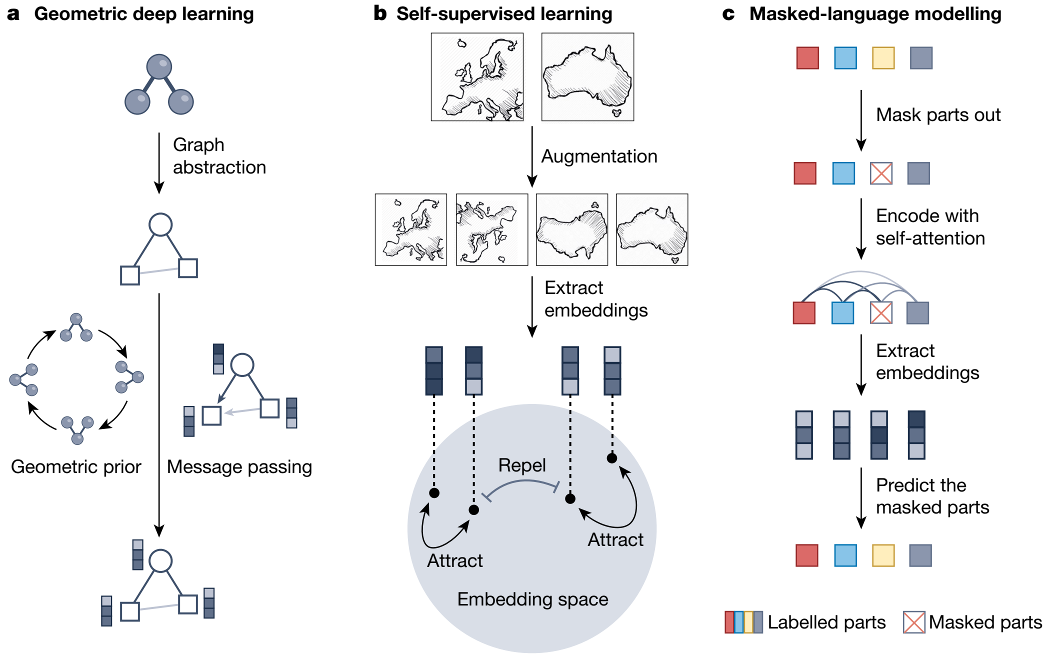

Fig. 2 | Learning meaningful representations of scientific data. **a**, Geometric deep learning integrates information about scientific data's geometry, structure and symmetry, such as molecules and materials, by leveraging graphs and employing neural message-passing strategies. This approach generates latent representations (embeddings) by exchanging neural messages along edges within graphs while considering other geometric priors, such as invariance and equivariance constraints. As a result, geometric deep learning can incorporate complex structural information into deep-learning models, allowing for a better understanding and manipulation of the underlying geometric datasets. **b**, To effectively represent diverse samples such as satellite images, it is crucial to capture both their similarities and differences. Self-supervised learning strategies, such as contrastive learning, achieve this by generating augmented counterparts and aligning positive while separating negative pairs. This iterative process enhances the embeddings, leading to informative latent representations and better performance on downstream prediction tasks. **c**, Masked-language modelling effectively captures the semantics of sequential data, such as natural language and biological sequences. This approach involves feeding masked elements of the input into a transformer block, which includes pre-processing steps, such as positional encodings. The self-attention mechanism, represented by grey lines with colour intensity reflecting the magnitude of attention weights, combines representations of non-masked input to accurately predict the masked input. This approach produces high-quality representations of sequences by repeating this autocompletion process across many elements of the input.

in masked-based training[99], the self-supervised task is to recover a masked token in a sequence using a bidirectional sequence context. Protein language models can encode amino acid sequences to capture structural and functional properties[100,101] and evaluate the evolutionary fitness of viral variants[102]. Such representations are transferable across various tasks, ranging from sequence design[103–105] to structure prediction[10,106]. In handling biochemical sequences[107–109], chemical language models facilitate efficient exploration of vast chemical space[110,111]. They have been used to predict properties[112], plan multi-step syntheses[113,114] and explore the space of chemical reactions[115–117].

Transformer architectures

Transformers (Box 1)[118] are neural architecture models that can process sequences of tokens by flexibly modelling interactions between arbitrary token pairs, surpassing earlier efforts using recurrent neural networks for sequential modelling. Transformers dominate natural language processing[37,99] and have been successfully applied to a range of problems, including earthquake signal detection[119], DNA and protein sequence modelling[10,120], modelling the effect of sequence variation on biological function[100,121], and symbolic regression[122]. Although transformers unify graph neural networks and language models[123–125], the run-time and memory footprint of transformers can scale quadratically with the length of sequences, leading to efficiency challenges addressed by long-range modelling[120] and linearized attention mechanisms[126]. As a result, unsupervised or self-supervised generative pre-trained transformers, followed by parameter-efficient fine-tuning, are widely used.

Neural operators

Standard neural network models can be inadequate for scientific applications as they assume a fixed data discretization. This approach is unsuitable for many scientific datasets collected at varying resolutions and grids. Moreover, data are often sampled from an underlying physical phenomenon in a continuous domain, such as seismic activity or fluid flow. Neural operators learn representations invariant to discretization by learning mappings between function spaces[127,128]. Neural operators are guaranteed to be discretization invariant, meaning that they can work on any discretization of inputs and converge to a limit upon mesh refinement. Once neural operators are trained, they can be evaluated at any resolution without the need for re-training. In contrast, the performance of standard neural networks can degrade when data resolution during deployment changes from model training.

AI-based generation of scientific hypotheses

Testable hypotheses are central to scientific discovery. They can take many forms, from symbolic expressions in mathematics to molecules in chemistry and genetic variants in biology. Formulating meaningful hypotheses can be a laborious process, as exemplified by Johannes Kepler, who spent four years analysing stellar and planetary data before arriving at a hypothesis that led to the discovery of the laws of planetary motion[129]. AI methods can be helpful at several stages of this process. They can generate hypotheses by identifying candidate symbolic expressions from noisy observations. They can help design objects, such as a molecule that binds to a therapeutic target[130] or a

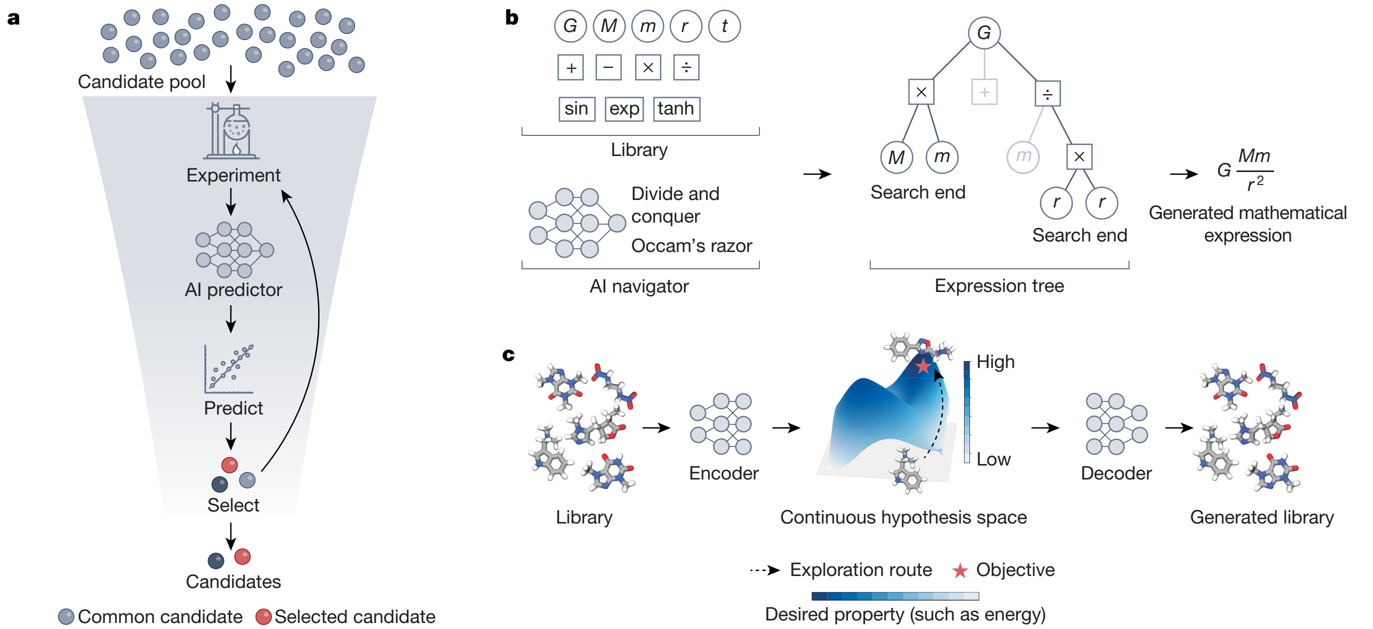

Fig. 3 | AI-guided generation of scientific hypotheses. a, High-throughput screening involves using AI predictors trained on experimentally generated datasets to select a small number of screened objects with desirable properties, thus reducing the size of the total candidate pool by orders of magnitude. This approach can leverage self-supervised learning to pre-train predictors on vast amounts of unscreened objects, followed by fine-tuning predictors on datasets of screened objects with labelled readouts. Laboratory evaluations and uncertainty quantification can refine this approach to streamline the screening process, making it more cost effective and time efficient, ultimately accelerating the identification of candidate chemical compounds, materials and biomolecules. **b**, The AI navigator employs rewards predicted by reinforcement-learning agents and design criteria, such as Occam's razor, to focus on the most promising elements of a candidate hypothesis during symbolic regression. Shown is an example illustrating the inference of the mathematical expression representing Newton's gravitational law. The low-scoring search routes are shown as grey branches in the symbolic expression tree. Guided by actions associated with the highest predicted rewards, this iterative process converges towards mathematical expressions consistent with the data and satisfying other design criteria. **c**, AI differentiators are autoencoder models that map discrete objects, such as chemical compounds, to points in a differentiable, continuous latent space. This space allows for the optimization of the objects, such as selecting compounds from a vast chemical library that maximize a specific biochemical endpoint. The idealized landscape plot depicts the learned latent space, with deeper colours indicating regions enriched for objects with higher predicted scores. By leveraging this latent space, the AI differentiator can efficiently identify objects that maximize the desired property indicated by the red star.

counterexample that contradicts a mathematical conjecture[9], suggesting experimental evaluation in the laboratory. Moreover, AI systems can learn a Bayesian posterior distribution (Box 1) of hypotheses and use it to generate hypotheses compatible with scientific data and knowledge[131].

Black-box predictors of scientific hypotheses

Identifying promising hypotheses for scientific enquiry requires efficiently examining many candidates and selecting those that can maximize the yield of downstream simulations and experimentation. In drug discovery, high-throughput screening can assess thousands to millions of molecules, and algorithms can prioritize which molecules to investigate experimentally[132]. Models can be trained to anticipate the utility of an experiment, such as relevant molecular properties[133,134] or symbolic formulas that fit the observations[122]. However, experimental ground-truth data for these predictors may be unavailable for many molecules. Thus, weak supervision-learning approaches (Box 1) can be used to train these models, where noisy, limited or imprecise supervision is used as a training signal. These serve as a cost-effective proxy for annotations from human experts, expensive in silico calculations or higher-fidelity experiments (Fig. 3a).

AI methods trained on high-fidelity simulations have been used to efficiently screen large libraries of molecules, such as 1.6 million organic-light-emitting-diode material candidates[133] and 11 billion synthon-based ligand candidates[134]. In genomics, transformer architectures trained to predict gene expression values from DNA sequences can help prioritize genetic variants[120]. In particle physics, identifying intrinsic charm quarks in protons involves screening all possible structures and fitting experimental data on each candidate structure[135]. To further increase the efficiency of these processes, AI-selected candidates can be sent to medium or low-throughput experiments for continual refinement of candidates using experimental feedback. The results can be fed back into the AI models using active learning[136] and Bayesian optimization[137] (Box 1), allowing the algorithms to refine their predictions and focus on the most promising candidates.

AI methods have become invaluable when hypotheses involve complex objects such as molecules. For instance, in protein folding, AlphaFold2[10] can predict the 3D atom coordinates of proteins from amino acid sequences with atomic accuracy, even for proteins whose structure is unlike any of the proteins in the training dataset. This breakthrough has led to the development of various AI-driven protein-folding methods, such as RoseTTAFold[106]. In addition to forward problems, AI approaches are increasingly used for inverse problems that aim to understand the causal factors that produced a set of observations. Inverse problems, such as inverse folding or fixed backbone design, can predict the amino acid sequence from the protein's backbone 3D atom coordinates using a black-box predictor trained on millions of protein structures[105]. However, such black-box AI predictors require large training datasets and offer limited interpretability despite reducing the dependence on the availability of prior scientific knowledge.

Navigating combinatorial hypothesis spaces

Although sampling all the hypotheses compatible with the data is daunting, a manageable goal is to search for a single good one, which can be formulated as an optimization problem. Instead of traditional methods that rely on manually engineered rules[138], AI policies can be used to estimate the reward of each search and prioritize search directions with higher values. An agent trained by a reinforcement-learning algorithm is typically employed to learn the policy. The agent learns to take actions in the search space that maximize a reward signal, which can be defined to reflect the quality of the generated hypotheses or other relevant criteria.

To solve the optimization problem, a symbolic regression task can be solved using evolutionary algorithms, which generate random symbolic laws as the initial set of solutions. Within each generation, slight variations are imposed on candidate solutions. The algorithm checks whether any modification produced a symbolic law that fits the observations better than prior solutions, keeping the best ones for the next generation[139]. However, reinforcement-learning approaches are increasingly replacing this standard strategy. Reinforcement learning uses neural networks to generate a mathematical expression sequentially by adding mathematical symbols from a predefined vocabulary and using the learned policy to decide which notation symbol to be added next[140]. The mathematical formula is represented as a parse tree. The learned policy takes the parse tree as input to determine what leaf node to expand and what notation (from the vocabulary) to add (Fig. 3b). Another approach for using neural networks to solve mathematical problems is transforming a mathematical formula into a binary sequence of symbols. A neural network policy can then probabilistically and sequentially grow the sequence one binary character at a time[6]. By designing a reward that measures the ability to refute the conjecture, this approach can find a refutation to a mathematical conjecture without prior knowledge about the mathematical problem.

Combinatorial optimization also applies to tasks such as discovering molecules with desirable pharmaceutical properties, where each step in molecular design is a discrete decision-making process. In this process, a partially generated molecular graph is given as input to the learned policy, making discrete choices on where to add a new atom and which atom to add at the selected position in the molecule. By iteratively performing this process, the policy can generate a series of possible molecular structures evaluated based on their fitness to the target properties. The search space is too vast to explore all possible combinations, but reinforcement learning can efficiently guide the search by prioritizing the most promising branches worth investigating[141–145]. Reinforcement-learning methods can be trained with a training objective that encourages the resulting policy to sample from all reasonable solutions (with a high reward) rather than to focus on a single good solution, as is the case with standard reward maximization in reinforcement learning[144–146]. These reinforcement-learning approaches have been successfully applied to various optimization problems, including maximizing protein expression[147], planning hydropower to reduce adverse impact in the Amazon Basin[148] and exploring the parameter space of particle accelerators[33].

Policies learned by AI agents have foresighted actions that seemed unconventional initially but proved to be effective[149]. For instance, in mathematics, supervised models can identify patterns and relations between mathematical objects and help guide intuition and propose conjectures[9]. These analyses have pointed to previously unknown patterns or even new models of the world. However, reinforcement-learning methods may not generalize well to unseen data during model training, as the agent may get stuck in a local optimum once it finds a sequence of actions that work well. To improve generalization, some exploration strategy is required to collect broader search trajectories that could help the agent perform better in new and modified settings.

Optimizing differentiable hypothesis spaces

Scientific hypotheses often take the form of discrete objects, such as symbolic formulas in physics or chemical compounds in pharmaceutical and materials science. Although combinatorial optimization techniques have been successful for some of these problems, a differentiable space can also be used for optimization as it is amenable to gradient-based methods, which can efficiently find local optima. To enable the use of gradient-based optimization, two approaches are frequently used. The first is to use models such as VAEs to map discrete candidate hypotheses to points in a latent differentiable space. The second approach is to relax discrete hypotheses into differentiable objects that can be optimized in the differentiable space. This relaxation can take different forms, such as replacing discrete variables with continuous ones or using a soft version of the original constraints.

Applications of symbolic regression in physics use grammar VAEs[150]. These models represent discrete symbolic expressions as parse trees using context-free grammar and map the trees into a differentiable latent space. Bayesian optimization is then employed to optimize the latent space for symbolic laws while ensuring that the expressions are syntactically valid. In a related study, Brunton and colleagues[151] introduced a method for differentiating symbolic rules by assigning trainable weights to predefined basis functions. Sparse regression was used to select a linear combination of the basis functions that accurately represented the dynamic system while maintaining compactness. Unlike equivariant neural networks, which use a predefined inductive bias to enforce symmetry, symmetry can be discovered as the characteristic behaviour of a domain. For instance, Liu and Tegmark[152] described asymmetry as a smooth loss function and minimized the loss function to extract previously unknown symmetries. This approach was applied to uncover hidden symmetries in black-hole waveform datasets, revealing unexpected space–time structures that were historically challenging to find.

In astrophysics, VAEs have been used to estimate gravitational-wave detector parameters based on pretrained black-hole waveform models. This method is up to six orders of magnitude faster than traditional methods, making it practical to capture transient gravitational-wave events[153]. In materials science, thermodynamic rules are combined with an autoencoder to design an interpretable latent space for identifying phase maps of crystal structures[154]. In chemistry, models such as simplified molecular-input line-entry system (SMILES)-VAE[155] can transform SMILES strings, which are molecular notations of chemical structures in the form of a discrete series of symbols that computers can easily understand, into a differentiable latent space that can be optimized using Bayesian optimization techniques (Fig. 3c). By representing molecular structures as points in the latent space, we can design differentiable objectives and optimize them using self-supervised learning to predict molecular properties based on latent representations of molecules. This means that we can optimize discrete molecular structures by backpropagating gradients of the AI predictor all the way to the continuous-valued representation of molecular inputs. A decoder can turn these molecular representations into approximately corresponding discrete inputs. This approach is used in the design of proteins[156] and small molecules[157,158].

Performing optimization in the latent space can more flexibly model underlying data distributions than mechanistic approaches in the original hypothesis space. However, extrapolative prediction in sparsely explored regions of the hypothesis space can be poor. In many scientific disciplines, hypothesis spaces can be vastly larger than what can be examined through experimentation. For instance, it is estimated that there are approximately 10^{60} molecules, whereas even the largest chemical libraries contain fewer than 10^{10} molecules[12,159]. Therefore, there is a pressing need for methods to efficiently search through and identify high-quality candidate solutions in these largely unexplored regions.

AI-driven experimentation and simulation

Evaluating scientific hypotheses through experimentation is critical to scientific discovery. However, laboratory experiments can be costly and

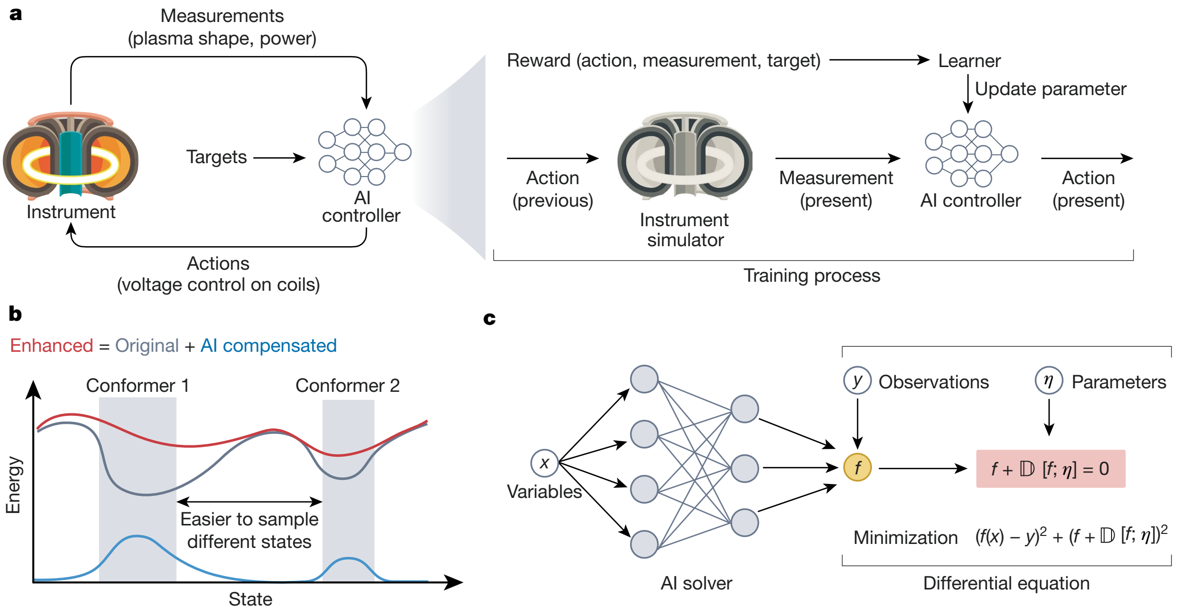

Fig. 4 | Integration of AI with scientific experiments and simulation. **a**, Leveraging AI for nuclear fusion control of complex and dynamic systems: Degrave et al.[166] developed an AI controller to regulate nuclear fusion through magnetic fields in a tokamak reactor. The AI agent receives real-time measurements of electrical voltage levels and plasma configurations and takes actions to control the magnetic field and meet experimental targets, such as maintaining a functional power supply. The controller is trained using simulations with a reward function to update model parameters. **b**, In computational simulations of complex systems, AI systems can accelerate the detection of rare events, such as transitions between different conformational structures of a protein. Wang et al.[169] used a neural-network-based uncertainty estimator to guide the addition of potentials that compensate for the original potential energy, allowing the system to escape local minima (in grey) and explore a configuration space more rapidly. This approach, illustrated here, can enhance the efficiency and accuracy of simulations, leading to a deeper understanding of complex biological phenomena. **c**, A neural framework for solving partial differential equations, where the AI solver is a physics-informed neural network trained to estimate target function *f*. The derivative of variable **x** is calculated by automatically differentiating the neural network's outputs. When the expression for the differential equation is unknown (parameterized by *η*), it can be estimated by solving a multi-objective loss that optimizes both the functional form of the equation and its fit to observations *y*. Credit: Nuclear fusion icon in **a**, iStockphoto/VectorMine.

impractical. Computer simulations have emerged as a promising alternative, offering the potential for more efficient and flexible experimentation. While simulations rely on handcrafted parameters and heuristics to imitate real-world scenarios, they require a trade-off between accuracy and speed compared with physical experimentation, necessitating understanding the underlying mechanisms. However, with the advent of deep learning, these challenges are being addressed by identifying and optimizing hypotheses for efficient testing and empowering computer simulations to link observations with hypotheses.

Efficient evaluation of scientific hypotheses

AI systems have provided experimental design and optimization tools, which can enhance traditional scientific methods, decrease the number of experiments needed and save resources. Specifically, AI systems can assist with two essential steps of experimental testing: planning and steering. In traditional approaches, these steps often require trial and error, which can be inefficient, costly and even life-threatening at times[160]. AI planning provides a systematic approach to designing experiments, optimizing their efficiency and exploring uncharted territory. At the same time, AI steering directs experimental processes towards high-yield hypotheses, allowing the system to learn from previous observations and adjust the course of experiments. These AI approaches can be model based, using simulations and prior knowledge, or model free, based on machine-learning algorithms alone.

AI systems can aid in the planning of experiments by optimizing the use of resources and reducing unnecessary investigations. Unlike hypothesis searching, experimental planning pertains to the procedures and steps involved in the design of scientific experiments. One example is synthesis planning in chemistry. Synthesis planning involves finding a sequence of steps by which a target chemical compound can be synthesized from available chemicals. AI systems can design synthetic routes to a desired chemical compound, reducing the need for human intervention[161,162]. Active learning has also been employed in materials discovery and synthesis[32,163–165]. Active learning involves iteratively interacting with and learning from experimental feedback to refine hypotheses. Material synthesis is a complex and resource-intensive process that requires efficient exploration of high-dimensional parameter space. Active learning uses uncertainty estimation to explore the parameter space and reduce uncertainty with as few steps as possible[165].

During an ongoing experiment, decision-making must often be adapted in real time. However, this process can be difficult and error prone when driven solely by human experience and intuition. Reinforcement learning provides an alternative approach that can continually react to the evolving environment and maximize the safety and success of the experiments. For example, reinforcement-learning approaches have proven to be effective for magnetic control of tokamak plasmas, where the algorithm interacts with the tokamak simulator to optimize a policy for controlling the process[166] (Fig. 4a). In another study, a reinforcement-learning agent used real-time feedback such as wind speed and solar elevation to control a stratospheric balloon and find favourable wind currents for navigation[14]. In quantum physics, experiment design needs to be dynamically adjusted as the best choice for a future materialization of a complex experiment can be counterintuitive. Reinforcement-learning methods can overcome this issue by iteratively designing the experiment and receiving feedback from it. For instance, reinforcement-learning algorithms have been used to optimize the measurement and control of quantum systems, where they improve experimental efficiency and accuracy[167].

Deducing observables from hypotheses using simulations

Computer simulation is a powerful tool to deduce observables from hypotheses, enabling the evaluation of hypotheses that are not directly testable. However, existing simulation techniques heavily rely on human understanding and knowledge about the underlying mechanisms of the studied systems, which can be suboptimal and inefficient. AI systems can enhance computer simulation with more accurate and efficient learning by better fitting key parameters of complex systems, solving differential equations that govern complex systems and modelling states in complex systems.

Scientists often study complex systems by creating a model that involves parameterized forms, which requires domain knowledge to identify initial symbolic expressions for the parameters. An example is molecular force fields, which are interpretable but limited in their ability to represent a wide range of functions and require strong inductive biases or scientific knowledge to generate. To improve the accuracy of molecular simulations, an AI-based neural potential that fits expensive yet accurate quantum-mechanical data has been developed to replace traditional force fields[16,168]. In addition, uncertainty quantification has been used to locate the energy barrier in the high-dimensional free-energy surface, thereby improving the efficiency of molecular dynamics[169] (Fig. 4b). For coarse-grained molecular dynamics, AI models have been utilized to reduce the computational cost for large systems by determining the degree to which the system needs to be coarsened from the learned hidden complex structures[170]. In quantum physics, neural networks have replaced manually estimated symbolic forms in parameterizing wave functions or density functionals due to their flexibility and ability to fit the data accurately[171,172].

Differential equations are crucial for modelling complex systems' dynamics in space and time. In contrast to numerical algebra solvers, AI-based neural solvers integrate data and physics more seamlessly[173,174]. These neural solvers combine physics with the flexibility of deep learning by grounding neural networks in domain knowledge (Fig. 4c). AI methods have been applied to solve differential equations in various fields, including computational fluid dynamics[175], predicting the structures of glassy systems[88], solving stiff chemical kinetic problems[176] and solving the Eikonal equation to characterize the travel times of seismic waves[177,178]. In dynamics modelling, continuous time can be modelled by neural ordinary differential equations[179]. Neural networks can parameterize solutions of Navier–Stokes equations in a spatiotemporal domain using physics-informed losses[180]. However, standard convolutional neural networks have limited ability to model fine-structured characteristics of the solution. This issue can be addressed by learning operators that model mappings between functions using neural networks[127,181]. In addition, solvers must be able to adapt to different domains and boundary conditions. This can be achieved by combining neural differential equations with graph neural networks to discretize arbitrary by graph partitioning[182].

Statistical modelling is a powerful tool to provide a full quantitative description of complex systems by modelling the distributions of states in those systems. Owing to its capability to capture highly complex distributions, deep generative modelling has recently emerged as a valuable approach in complex system simulations. One well known example is the Boltzmann generator[183] based on normalizing flows[184,185] (Box 1). Normalizing flows can map any complex distribution to a prior distribution (for example, a simple Gaussian) and back using a series of invertible neural networks. Although computationally expensive (often requiring hundreds or thousands of neural layers), normalizing flows provide an exact density function, which enables sampling and training. Unlike conventional simulations, normalizing flows can generate equilibrium states by directly sampling from the prior distribution and applying the neural network, which has a fixed computational cost. This enhances sampling in the lattice field[186] and gauge theories[187] and improves Markov chain Monte Carlo methods[188] that otherwise might not converge due to mode mixing[189–191].

Grand challenges

To harness scientific data, models must be built and employed with simulation and human expertise. Such integration has opened up opportunities for scientific discovery. However, to further enhance the impact of AI across scientific disciplines, significant progress is needed in theory, methods, software and hardware infrastructure. Cross-disciplinary collaborations are crucial to realize a comprehensive and practical approach towards advancing science through AI.

Practical considerations

Scientific datasets are often not directly amenable to AI analyses because of measurement technology limitations that produce incomplete datasets and biased or conflicting readouts, and limited accessibility owing to privacy and safety concerns. Standardized and transparent formats are required to alleviate the workload of data processing[159,192–196]. Model cards[197] and datasheets[198] are examples of efforts to document the operating characteristics of scientific datasets and models. In addition, federated learning[199,200] and cryptographic[201] algorithms can be used to prevent releasing sensitive data with high commercial value to the public domain. Leveraging open scientific literature, natural language processing and knowledge graph techniques can facilitate literature mining to support material discovery[15], chemistry synthesis[202] and therapeutic science[203].

The use of deep learning poses a complex challenge for human-in-the-loop AI-driven design, discovery and evaluation. To automate scientific workflows, optimize large-scale simulation codes and operate instruments, autonomous robotic control can leverage predictions and conduct experiments on high-throughput synthesis and testing lines, creating self-driving laboratories. The early application of generative models in materials exploration suggests that millions of possible materials could be identified with desired properties and functions and evaluated for synthesizability. For instance, King et al.[204] combined logical AI and robotics to autonomously generate functional genomics hypotheses about yeast and experimentally test the hypotheses using laboratory automation. In chemical synthesis, AI optimizes candidate synthesis routes, followed by robots steering chemical reactions in predicted synthesis routes[7].

The practical implementation of an AI system involves complex software and hardware engineering, requiring a series of interdependent steps that go from data curation and processing to algorithm implementation and design of user and application interfaces. Minor variations in implementation can lead to considerable changes in performance and impact the success of integrating AI models within scientific practice. Therefore, both data and model standardization needs to be considered. AI approaches can suffer from reproducibility due to the stochastic nature of model training, varying model parameters and evolving training datasets, which are both data dependent and task dependent. Standardized benchmarks and experimental design can alleviate such issues[205]. Another direction towards improving reproducibility is through open-source initiatives that release open models, datasets and education programmes[4,130,206,207].

Algorithmic innovations

To contribute to scientific understanding or acquire it autonomously, algorithmic innovation is required to establish a foundational ecosystem with the most appropriate algorithms for use throughout the scientific process.

The question of out-of-distribution generalization is at the frontier of AI research. A neural network trained on data from a specific regime may discover regularities that do not generalize in a different regime whose underlying distribution has shifted (Box 1). Although many scientific laws are not universal, their applicability is generally broad. Compared with state-of-the-art AI, human brains can better and faster generalize to modified settings. An attractive hypothesis is that this is because humans build not just a statistical model of what they observe but a causal model, that is, a family of statistical models indexed by all possible interventions (for example, different initial states, actions of agents or different regimes). Incorporating causality in AI is still a young field[208–212] where much remains to be done. Techniques such as self-supervised learning have great potential for scientific problems because they can leverage massive unlabelled data and transfer their knowledge to low-data regimes. However, current transfer-learning schemes can be ad hoc, lack theoretical guidance[213] and are vulnerable to shifts in underlying distributions[214]. Although preliminary attempts have addressed this challenge[215,216], more exploration is needed to systematically measure transferability across domains and prevent negative transfer. Moreover, to address the difficulties that scientists care about, the development and evaluation of AI methods must be done in real-world scenarios, such as plausibly realizable synthesis paths in drug design[217,218], and include well calibrated uncertainty estimators to assess the model's reliability before transitioning it to real-world implementation.

Scientific data are multimodal and include images (such as black-hole images in cosmology), natural language (such as scientific literature), time series (such as thermal yellowing of materials), sequences (such as biological sequences), graphs (such as complex systems) and structures (such as 3D protein–ligand conformations). For instance, in high-energy physics, jets are collimated sprays of particles produced from quarks and gluons at high energy. Identifying their substructures from radiation patterns can aid in the search for new physics. The jet substructures can be described by images, sequences, binary trees, generic graphs and sets of tensors[18]. Although using neural networks to process images has been extensively researched, processing particle images alone is insufficient. Similarly, using other representations of jet substructures in isolation cannot give a holistic and integrated systems view of the complex system[219]. Although integrating multimodal observations remains a challenge, the modular nature of neural networks implies that distinct neural modules can transform diverse data modalities into universal vector representations[220,221].

Scientific knowledge, such as rotational equivariance in molecules[77], equality constraints in mathematics[182], disease mechanisms in biology[222] and multi-scale structures in complex systems[223,224], can be incorporated into AI models. However, which principles and knowledge are most helpful and practical to implement is still unclear. As AI models require massive data to fit, incorporating scientific knowledge into models can aid learning when datasets are small or sparsely annotated. Therefore, research must establish principled methods for integrating knowledge into AI models and understanding the trade-offs between domain knowledge and learning from measured data.

AI methods often operate as black boxes, meaning that users cannot fully explain how outputs have been generated and what inputs have been critical in producing the outputs. Black-box models can decrease user trust in predictions and have limited applicability in areas where model outputs must be understood before real-world implementation[225–227], such as in human space exploration[228], and where predictions inform policy, such as in climate science[229]. Transparent deep-learning models remain elusive[230] despite a plethora of explainability techniques[231–233]. However, the fact that human brains can synthesize high-level explanations, even if imperfect, that can convince other humans offers hope that by modelling phenomena at similarly high levels of abstraction, future AI models will provide interpretable explanations at least as valuable as those offered by human brains. This also suggests that studying higher-level cognition could potentially inspire future deep-learning models to incorporate both current deep-learning abilities and the abilities to manipulate verbalizable abstractions, reason causally, and generalize out of distribution.

Conduct of science and scientific enterprise

Looking towards the future, the demand for AI expertise will be influenced by two forces. First, the existence of problems that are on the verge of benefiting from the application of AI—such as self-driving labs. Second, the ability of intelligent tools to enhance the state-of-the-art and create new opportunities—such as examining biological, chemical or physical processes that happen at length and time scales not accessible in experiments. On the basis of these two forces, we anticipate that research teams will change in composition to include AI specialists, software and hardware engineers, and novel forms of collaboration involving government at all levels, educational institutions and corporations. Recent state-of-the-art deep-learning models continue to grow in size[10,234]. These models consist of millions or even billions of parameters and have experienced a tenfold increase in size year on year. Training these models involves transmitting data through complex parameterized mathematical operations, with parameters updated to nudge the model outputs towards the desired values. However, the computational and data requirements to calculate these updates are colossal, resulting in a large energy footprint and high computational costs. As a result, big tech companies have heavily invested in computational infrastructure and cloud services, pushing the limits on scale and efficiency. Although for-profit and non-academic organizations have access to vast computational infrastructure, higher-education institutions can be better integrated across multiple disciplines. Furthermore, academic institutions tend to host unique historical databases and measurement technology that might not exist elsewhere but are necessary for AI4Science. These complementary assets have facilitated new modes of industry–academia partnerships, which can impact the selection of research questions pursued.

As AI systems approach performance that rivals and surpasses humans, employing it as a drop-in replacement for routine laboratory work is becoming feasible. This approach enables researchers to develop predictive models from experimental data iteratively and select experiments to improve them without manually performing laborious and repetitive tasks[217,235]. To support this paradigm shift, educational programmes are emerging to train scientists in designing, implementing and applying laboratory automation and AI in scientific research. These programmes help scientists understand when the use of AI is appropriate and to prevent misinterpreted conclusions from AI analyses.

The misapplication of AI tools and misinterpretation of their results can have significant negative impacts[236]. The broad range of applications compounds these risks[237]. However, the misuse of AI is not solely a technological problem; it also depends on the incentives of those leading AI innovation and investing in AI implementation. Establishing ethics review processes and responsible implementation tactics is essential, including a comprehensive overview of the scope and applicability of AI[238]. Furthermore, security risks associated with AI must be considered, as it has become easier to repurpose algorithm implementations for dual use[237]. As algorithms are adaptable to a broad range of applications, they can be developed for one purpose but used for another, creating vulnerabilities to threats and manipulation.

Conclusion

AI systems can contribute to scientific understanding, enable the investigation of processes and objects that cannot be visualized or probed in any other way, and systematically inspire ideas by building models from data and combining them with simulation and scalable computing. To realize this potential, safety and security concerns that come with the use of AI must be addressed through responsible and thoughtful deployment of the technology. To use AI responsibly in scientific research, we need to measure the levels of uncertainty, error, and utility of AI systems. This understanding is essential for accurately interpreting AI outputs and ensuring that we do not rely too heavily on potentially flawed results. As AI systems continue to evolve, prioritizing reliable implementation with proper safeguards in place is key to minimizing risks and maximizing benefits. AI has the potential to unlock scientific discoveries that were previously out of reach.

1. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
   This survey summarizes key elements of deep learning and its development in speech recognition, computer vision and and natural language processing.
2. de Regt, H. W. Understanding, values, and the aims of science. *Phil. Sci.* **87**, 921–932 (2020).
3. Pickstone, J. V. *Ways of Knowing: A New History of Science, Technology, and Medicine* (Univ. Chicago Press, 2001).
4. Han, J. et al. Deep potential: a general representation of a many-body potential energy surface. *Commun. Comput. Phys.* **23**, 629–639 (2018).
   This paper introduced a deep neural network architecture that learns the potential energy surface of many-body systems while respecting the underlying symmetries of the system by incorporating group theory.
5. Akiyama, K. et al. First M87 Event Horizon Telescope results. IV. Imaging the central supermassive black hole. *Astrophys. J. Lett.* **875**, L4 (2019).
6. Wagner, A. Z. Constructions in combinatorics via neural networks. Preprint at https://arxiv.org/abs/2104.14516 (2021).
7. Coley, C. W. et al. A robotic platform for flow synthesis of organic compounds informed by AI planning. *Science* **365**, eaax1566 (2019).
8. Bommasani, R. et al. On the opportunities and risks of foundation models. Preprint at https://arxiv.org/abs/2108.07258 (2021).
9. Davies, A. et al. Advancing mathematics by guiding human intuition with AI. *Nature* **600**, 70–74 (2021).
   This paper explores how AI can aid the development of pure mathematics by guiding mathematical intuition.
10. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
    This study was the first to demonstrate the ability to predict protein folding structures using AI methods with a high degree of accuracy, achieving results that are at or near the experimental resolution. This accomplishment is particularly noteworthy, as predicting protein folding has been a grand challenge in the field of molecular biology for over 50 years.
11. Stokes, J. M. et al. A deep learning approach to antibiotic discovery. *Cell* **180**, 688–702 (2020).
12. Bohacek, R. S., McMartin, C. & Guida, W. C. The art and practice of structure-based drug design: a molecular modeling perspective. *Med. Res. Rev.* **16**, 3–50 (1996).
13. Bileschi, M. L. et al. Using deep learning to annotate the protein universe. *Nat. Biotechnol.* **40**, 932–937 (2022).
14. Bellemare, M. G. et al. Autonomous navigation of stratospheric balloons using reinforcement learning. *Nature* **588**, 77–82 (2020).
    This paper describes a reinforcement-learning algorithm for navigating a super-pressure balloon in the stratosphere, making real-time decisions in the changing environment.
15. Tshitoyan, V. et al. Unsupervised word embeddings capture latent knowledge from materials science literature. *Nature* **571**, 95–98 (2019).
16. Zhang, L. et al. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Phys. Rev. Lett.* **120**, 143001 (2018).
17. Deiana, A. M. et al. Applications and techniques for fast machine learning in science. *Front. Big Data* **5**, 787421 (2022).
18. Karagiorgi, G. et al. Machine learning in the search for new fundamental physics. *Nat. Rev. Phys.* **4**, 399–412 (2022).
19. Zhou, C. & Paffenroth, R. C. Anomaly detection with robust deep autoencoders. In *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 665–674 (2017).
20. Hinton, G. E. & Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science* **313**, 504–507 (2006).
21. Kasieczka, G. et al. The LHC Olympics 2020 a community challenge for anomaly detection in high energy physics. *Rep. Prog. Phys.* **84**, 124201 (2021).
22. Govorkova, E. et al. Autoencoders on field-programmable gate arrays for real-time, unsupervised new physics detection at 40 MHz at the Large Hadron Collider. *Nat. Mach. Intell.* **4**, 154–161 (2022).
23. Chamberland, M. et al. Detecting microstructural deviations in individuals with deep diffusion MRI tractometry. *Nat. Comput. Sci.* **1**, 598–606 (2021).
24. Rafique, M. et al. Delegated regressor, a robust approach for automated anomaly detection in the soil radon time series data. *Sci. Rep.* **10**, 3004 (2020).
25. Pastore, V. P. et al. Annotation-free learning of plankton for classification and anomaly detection. *Sci. Rep.* **10**, 12142 (2020).
26. Naul, B. et al. A recurrent neural network for classification of unevenly sampled variable stars. *Nat. Astron.* **2**, 151–155 (2018).
27. Lee, D.-H. et al. Pseudo-label: the simple and efficient semi-supervised learning method for deep neural networks. In *ICML Workshop on Challenges in Representation Learning* (2013).
28. Zhou, D. et al. Learning with local and global consistency. In *Advances in Neural Information Processing Systems* **16**, 321–328 (2003).
29. Radivojac, P. et al. A large-scale evaluation of computational protein function prediction. *Nat. Methods* **10**, 221–227 (2013).
30. Barkas, N. et al. Joint analysis of heterogeneous single-cell RNA-seq dataset collections. *Nat. Methods* **16**, 695–698 (2019).
31. Tran, K. & Ulissi, Z. W. Active learning across intermetallics to guide discovery of electrocatalysts for CO_2 reduction and H_2 evolution. *Nat. Catal.* **1**, 696–703 (2018).
32. Jablonka, K. M. et al. Bias free multiobjective active learning for materials design and discovery. *Nat. Commun.* **12**, 2312 (2021).
33. Roussel, R. et al. Turn-key constrained parameter space exploration for particle accelerators using Bayesian active learning. *Nat. Commun.* **12**, 5612 (2021).
34. Ratner, A. J. et al. Data programming: creating large training sets, quickly. In *Advances in Neural Information Processing Systems* **29**, 3567–3575 (2016).
35. Ratner, A. et al. Snorkel: rapid training data creation with weak supervision. In *International Conference on Very Large Data Bases* **11**, 269–282 (2017).
    This paper presents a weakly-supervised AI system designed to annotate massive amounts of data using labeling functions.
36. Butter, A. et al. GANplifying event samples. *SciPost Phys.* **10**, 139 (2021).
37. Brown, T. et al. Language models are few-shot learners. In *Advances in Neural Information Processing Systems* **33**, 1877–1901 (2020).
38. Ramesh, A. et al. Zero-shot text-to-image generation. In *International Conference on Machine Learning* **139**, 8821–8831 (2021).
39. Littman, M. L. Reinforcement learning improves behaviour from evaluative feedback. *Nature* **521**, 445–451 (2015).
40. Cubuk, E. D. et al. Autoaugment: learning augmentation strategies from data. In *IEEE Conference on Computer Vision and Pattern Recognition* 113–123 (2019).
41. Reed, C. J. et al. Selfaugment: automatic augmentation policies for self-supervised learning. In *IEEE Conference on Computer Vision and Pattern Recognition* 2674–2683 (2021).
42. ATLAS Collaboration et al. Deep generative models for fast photon shower simulation in ATLAS. Preprint at https://arxiv.org/abs/2210.06204 (2022).
43. Mahmood, F. et al. Deep adversarial training for multi-organ nuclei segmentation in histopathology images. *IEEE Trans. Med. Imaging* **39**, 3257–3267 (2019).
44. Teixeira, B. et al. Generating synthetic X-ray images of a person from the surface geometry. In *IEEE Conference on Computer Vision and Pattern Recognition* 9059–9067 (2018).
45. Lee, D., Moon, W.-J. & Ye, J. C. Assessing the importance of magnetic resonance contrasts using collaborative generative adversarial networks. *Nat. Mach. Intell.* **2**, 34–42 (2020).
46. Kench, S. & Cooper, S. J. Generating three-dimensional structures from a two-dimensional slice with generative adversarial network-based dimensionality expansion. *Nat. Mach. Intell.* **3**, 299–305 (2021).
47. Wan, C. & Jones, D. T. Protein function prediction is improved by creating synthetic feature samples with generative adversarial networks. *Nat. Mach. Intell.* **2**, 540–550 (2020).
48. Repecka, D. et al. Expanding functional protein sequence spaces using generative adversarial networks. *Nat. Mach. Intell.* **3**, 324–333 (2021).
49. Marouf, M. et al. Realistic in silico generation and augmentation of single-cell RNA-seq data using generative adversarial networks. *Nat. Commun.* **11**, 166 (2020).
50. Ghahramani, Z. Probabilistic machine learning and artificial intelligence. *Nature* **521**, 452–459 (2015).
    This survey provides an introduction to probabilistic machine learning, which involves the representation and manipulation of uncertainty in models and predictions, playing a central role in scientific data analysis.
51. Cogan, J. et al. Jet-images: computer vision inspired techniques for jet tagging. *J. High Energy Phys.* **2015**, 118 (2015).
52. Zhao, W. et al. Sparse deconvolution improves the resolution of live-cell super-resolution fluorescence microscopy. *Nat. Biotechnol.* **40**, 606–617 (2022).
53. Brbić, M. et al. MARS: discovering novel cell types across heterogeneous single-cell experiments. *Nat. Methods* **17**, 1200–1206 (2020).
54. Qiao, C. et al. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nat. Methods* **18**, 194–202 (2021).
55. Andreassen, A. et al. OmniFold: a method to simultaneously unfold all observables. *Phys. Rev. Lett.* **124**, 182001 (2020).
56. Bergenstråhle, L. et al. Super-resolved spatial transcriptomics by deep data fusion. *Nat. Biotechnol.* **40**, 476–479 (2021).
57. Vincent, P. et al. Extracting and composing robust features with denoising autoencoders. In *International Conference on Machine Learning* 1096–1103 (2008).
58. Kingma, D. P. & Welling, M. Auto-encoding variational Bayes. In *International Conference on Learning Representations* (2014).
59. Eraslan, G. et al. Single-cell RNA-seq denoising using a deep count autoencoder. *Nat. Commun.* **10**, 390 (2019).
60. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (MIT Press, 2016).
61. Olshausen, B. A. & Field, D. J. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* **381**, 607–609 (1996).
62. Bengio, Y. Deep learning of representations for unsupervised and transfer learning. In *ICML Workshop on Unsupervised and Transfer Learning* (2012).
63. Detlefsen, N. S., Hauberg, S. & Boomsma, W. Learning meaningful representations of protein sequences. *Nat. Commun.* **13**, 1914 (2022).
64. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2019).
65. Bronstein, M. M. et al. Geometric deep learning: going beyond euclidean data. *IEEE Signal Process Mag.* **34**, 18–42 (2017).
66. Anderson, P. W. More is different: broken symmetry and the nature of the hierarchical structure of science. *Science* **177**, 393–396 (1972).
67. Qiao, Z. et al. Informing geometric deep learning with electronic interactions to accelerate quantum chemistry. *Proc. Natl Acad. Sci. USA* **119**, e2205221119 (2022).
68. Bogatskiy, A. et al. Symmetry group equivariant architectures for physics. Preprint at https://arxiv.org/abs/2203.06153 (2022).
69. Bronstein, M. M. et al. Geometric deep learning: grids, groups, graphs, geodesics, and gauges. Preprint at https://arxiv.org/abs/2104.13478 (2021).

70. Townshend, R. J. L. et al. Geometric deep learning of RNA structure. *Science* **373**, 1047–1051 (2021).
71. Wicky, B. I. M. et al. Hallucinating symmetric protein assemblies. *Science* **378**, 56–61 (2022).
72. Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations* (2017).
73. Veličković, P. et al. Graph attention networks. In *International Conference on Learning Representations* (2018).
74. Hamilton, W. L., Ying, Z. & Leskovec, J. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems* **30**, 1024–1034 (2017).
75. Gilmer, J. et al. Neural message passing for quantum chemistry. In *International Conference on Machine Learning* 1263–1272 (2017).
76. Li, M. M., Huang, K. & Zitnik, M. Graph representation learning in biomedicine and healthcare. *Nat. Biomed. Eng.* **6**, 1353–1369 (2022).
77. Satorras, V. G., Hoogeboom, E. & Welling, M. E(*n*) equivariant graph neural networks. In *International Conference on Machine Learning* 9323–9332 (2021).
This study incorporates principles of physics into the design of neural models, advancing the field of equivariant machine learning.
78. Thomas, N. et al. Tensor field networks: rotation-and translation-equivariant neural networks for 3D point clouds. Preprint at https://arxiv.org/abs/1802.08219 (2018).
79. Finzi, M. et al. Generalizing convolutional neural networks for equivariance to lie groups on arbitrary continuous data. In *International Conference on Machine Learning* 3165–3176 (2020).
80. Fuchs, F. et al. SE(3)-transformers: 3D roto-translation equivariant attention networks. In *Advances in Neural Information Processing Systems* **33**, 1970-1981 (2020).
81. Zaheer, M. et al. Deep sets. In *Advances in Neural Information Processing Systems* **30**, 3391–3401 (2017).
This paper is an early study that explores the use of deep neural architectures on set data, which consists of an unordered list of elements.
82. Cohen, T. S. et al. Spherical CNNs. In *International Conference on Learning Representations* (2018).
83. Gordon, J. et al. Permutation equivariant models for compositional generalization in language. In *International Conference on Learning Representations* (2019).
84. Finzi, M., Welling, M. & Wilson, A. G. A practical method for constructing equivariant multilayer perceptrons for arbitrary matrix groups. In *International Conference on Machine Learning* 3318–3328 (2021).
85. Dijk, D. V. et al. Recovering gene interactions from single-cell data using data diffusion. *Cell* **174**, 716–729 (2018).
86. Gainza, P. et al. Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat. Methods* **17**, 184–192 (2020).
87. Hatfield, P. W. et al. The data-driven future of high-energy-density physics. *Nature* **593**, 351–361 (2021).
88. Bapst, V. et al. Unveiling the predictive power of static structure in glassy systems. *Nat. Phys.* **16**, 448–454 (2020).
89. Zhang, R., Zhou, T. & Ma, J. Multiscale and integrative single-cell Hi-C analysis with Higashi. *Nat. Biotechnol.* **40**, 254–261 (2022).
90. Sammut, S.-J. et al. Multi-omic machine learning predictor of breast cancer therapy response. *Nature* **601**, 623–629 (2022).
91. DeZoort, G. et al. Graph neural networks at the Large Hadron Collider. *Nat. Rev. Phys.* **5**, 281–303 (2023).
92. Liu, S. et al. Pre-training molecular graph representation with 3D geometry. In *International Conference on Learning Representations* (2022).
93. The LIGO Scientific Collaboration. et al. A gravitational-wave standard siren measurement of the Hubble constant. *Nature* **551**, 85–88 (2017).
94. Reichstein, M. et al. Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204 (2019).
95. Goenka, S. D. et al. Accelerated identification of disease-causing variants with ultra-rapid nanopore genome sequencing. *Nat. Biotechnol.* **40**, 1035–1041 (2022).
96. Bengio, Y. et al. Greedy layer-wise training of deep networks. In *Advances in Neural Information Processing Systems* **19**, 153–160 (2006).
97. Hinton, G. E., Osindero, S. & Teh, Y.-W. A fast learning algorithm for deep belief nets. *Neural Comput.* **18**, 1527–1554 (2006).
98. Jordan, M. I. & Mitchell, T. M. Machine learning: trends, perspectives, and prospects. *Science* **349**, 255–260 (2015).
99. Devlin, J. et al. BERT: pre-training of deep bidirectional transformers for language understanding. In *North American Chapter of the Association for Computational Linguistics* 4171–4186 (2019).
100. Rives, A. et al. Biological structure and function emerge from scaling unsupervised learning to 250 million protein sequences. *Proc. Natl Acad. Sci. USA* **118**, e2016239118 (2021).
101. Elnaggar, A. et al. ProtTrans: rowards cracking the language of lifes code through self-supervised deep learning and high performance computing. In *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).
102. Hie, B. et al. Learning the language of viral evolution and escape. *Science* **371**, 284–288 (2021).
This paper modeled viral escape with machine learning algorithms originally developed for human natural language.
103. Biswas, S. et al. Low-*N* protein engineering with data-efficient deep learning. *Nat. Methods* **18**, 389–396 (2021).
104. Ferruz, N. & Höcker, B. Controllable protein design with language models. *Nat. Mach. Intell.* **4**, 521–532 (2022).
105. Hsu, C. et al. Learning inverse folding from millions of predicted structures. In *International Conference on Machine Learning* 8946–8970 (2022).
106. Baek, M. et al. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–876 (2021).
Inspired by AlphaFold2, this study reported RoseTTAFold, a novel three-track neural module capable of simultaneously processing protein's sequence, distance and coordinates.
107. Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.* **28**, 31–36 (1988).
108. Lin, T.-S. et al. BigSMILES: a structurally-based line notation for describing macromolecules. *ACS Cent. Sci.* **5**, 1523–1531 (2019).
109. Krenn, M. et al. SELFIES and the future of molecular string representations. *Patterns* **3**, 100588 (2022).
110. Flam-Shepherd, D., Zhu, K. & Aspuru-Guzik, A. Language models can learn complex molecular distributions. *Nat. Commun.* **13**, 3293 (2022).
111. Skinnider, M. A. et al. Chemical language models enable navigation in sparsely populated chemical space. *Nat. Mach. Intell.* **3**, 759–770 (2021).
112. Chithrananda, S., Grand, G. & Ramsundar, B. ChemBERTa: large-scale self-supervised pretraining for molecular property prediction. In *Machine Learning for Molecules Workshop at NeurIPS* (2020).
113. Schwaller, P. et al. Predicting retrosynthetic pathways using transformer-based models and a hyper-graph exploration strategy. *Chem. Sci.* **11**, 3316–3325 (2020).
114. Tetko, I. V. et al. State-of-the-art augmented NLP transformer models for direct and single-step retrosynthesis. *Nat. Commun.* **11**, 5575 (2020).
115. Schwaller, P. et al. Mapping the space of chemical reactions using attention-based neural networks. *Nat. Mach. Intell.* **3**, 144–152 (2021).
116. Kovács, D. P., McCorkindale, W. & Lee, A. A. Quantitative interpretation explains machine learning models for chemical reaction prediction and uncovers bias. *Nat. Commun.* **12**, 1695 (2021).
117. Pesciullesi, G. et al. Transfer learning enables the molecular transformer to predict regio-and stereoselective reactions on carbohydrates. *Nat. Commun.* **11**, 4874 (2020).
118. Vaswani, A. et al. Attention is all you need. In *Advances in Neural Information Processing Systems* **30**, 5998–6008 (2017).
This paper introduced the transformer, a modern neural network architecture that can process sequential data in parallel, revolutionizing natural language processing and sequence modeling.
119. Mousavi, S. M. et al. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nat. Commun.* **11**, 3952 (2020).
120. Avsec, Ž. et al. Effective gene expression prediction from sequence by integrating long-range interactions. *Nat. Methods* **18**, 1196–1203 (2021).
121. Meier, J. et al. Language models enable zero-shot prediction of the effects of mutations on protein function. In *Advances in Neural Information Processing Systems* **34**, 29287–29303 (2021).
122. Kamienny, P.-A. et al. End-to-end symbolic regression with transformers. In *Advances in Neural Information Processing Systems* **35**, 10269–10281 (2022).
123. Jaegle, A. et al. Perceiver: general perception with iterative attention. In *International Conference on Machine Learning* 4651–4664 (2021).
124. Chen, L. et al. Decision transformer: reinforcement learning via sequence modeling. In *Advances in Neural Information Processing Systems* **34**, 15084–15097 (2021).
125. Dosovitskiy, A. et al. An image is worth 16x16 words: transformers for image recognition at scale. In *International Conference on Learning Representations* (2020).
126. Choromanski, K. et al. Rethinking attention with performers. In *International Conference on Learning Representations* (2021).
127. Li, Z. et al. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations* (2021).
128. Kovachki, N. et al. Neural operator: learning maps between function spaces. *J. Mach. Learn. Res.* **24**, 1–97 (2023).
129. Russell, J. L. Kepler's laws of planetary motion: 1609–1666. *Br. J. Hist. Sci.* **2**, 1–24 (1964).
130. Huang, K. et al. Artificial intelligence foundation for therapeutic science. *Nat. Chem. Biol.* **18**, 1033–1036 (2022).
131. Guimerà, R. et al. A Bayesian machine scientist to aid in the solution of challenging scientific problems. *Sci. Adv.* **6**, eaav6971 (2020).
132. Liu, G. et al. Deep learning-guided discovery of an antibiotic targeting Acinetobacter baumannii. *Nat. Chem. Biol.* https://doi.org/10.1038/s41589-023-01349-8 (2023).
133. Gómez-Bombarelli, R. et al. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nat. Mater.* **15**, 1120–1127 (2016).
This paper proposes using a black-box AI predictor to accelerate high-throughput screening of molecules in materials science.
134. Sadybekov, A. A. et al. Synthon-based ligand discovery in virtual libraries of over 11 billion compounds. *Nature* **601**, 452–459 (2022).
135. The NNPDF Collaboration Evidence for intrinsic charm quarks in the proton. *Nature* **606**, 483–487 (2022).
136. Graff, D. E., Shakhnovich, E. I. & Coley, C. W. Accelerating high-throughput virtual screening through molecular pool-based active learning. *Chem. Sci.* **12**, 7866–7881 (2021).
137. Janet, J. P. et al. Accurate multiobjective design in a space of millions of transition metal complexes with neural-network-driven efficient global optimization. *ACS Cent. Sci.* **6**, 513–524 (2020).
138. Bacon, F. *Novum Organon* Vol. 1620 (2000).
139. Schmidt, M. & Lipson, H. Distilling free-form natural laws from experimental data. *Science* **324**, 81–85 (2009).
140. Petersen, B. K. et al. Deep symbolic regression: recovering mathematical expressions from data via risk-seeking policy gradients. In *International Conference on Learning Representations* (2020).
141. Zhavoronkov, A. et al. Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* **37**, 1038–1040 (2019).
This paper describes a reinforcement-learning algorithm for navigating molecular combinatorial spaces, and it validates generated molecules using wet-lab experiments.
142. Zhou, Z. et al. Optimization of molecules via deep reinforcement learning. *Sci. Rep.* **9**, 10752 (2019).
143. You, J. et al. Graph convolutional policy network for goal-directed molecular graph generation. In *Advances in Neural Information Processing Systems* **31**, 6412–6422 (2018).

144. Bengio, Y. et al. GFlowNet foundations. Preprint at https://arxiv.org/abs/2111.09266 (2021).
This paper describes a generative flow network that generates objects by sampling them from a distribution optimized for drug design.
145. Jain, M. et al. Biological sequence design with GFlowNets. In *International Conference on Machine Learning* 9786–9801 (2022).
146. Malkin, N. et al. Trajectory balance: improved credit assignment in GFlowNets. In *Advances in Neural Information Processing Systems* **35**, 5955–5967 (2022).
147. Borkowski, O. et al. Large scale active-learning-guided exploration for in vitro protein production optimization. *Nat. Commun.* **11**, 1872 (2020).
This study introduced a dynamic programming approach to determine the optimal locations and capacities of hydropower dams in the Amazon Basin, balancing between energy production and environmental impact.
148. Flecker, A. S. et al. Reducing adverse impacts of Amazon hydropower expansion. *Science* **375**, 753–760 (2022).
This study introduced a dynamic programming approach to determine the optimal locations and capacities of hydropower dams in the Amazon basin, achieving a balance between the benefits of energy production and the potential environmental impacts.
149. Pion-Tonachini, L. et al. Learning from learning machines: a new generation of AI technology to meet the needs of science. Preprint at https://arxiv.org/abs/2111.13786 (2021).
150. Kusner, M. J., Paige, B. & Hernández-Lobato, J. M. Grammar variational autoencoder. In *International Conference on Machine Learning* 1945–1954 (2017).
This paper describes a grammar variational autoencoder that generates novel symbolic laws and drug molecules.
151. Brunton, S. L., Proctor, J. L. & Kutz, J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA* **113**, 3932–3937 (2016).
152. Liu, Z. & Tegmark, M. Machine learning hidden symmetries. *Phys. Rev. Lett.* **128**, 180201 (2022).
153. Gabbard, H. et al. Bayesian parameter estimation using conditional variational autoencoders for gravitational-wave astronomy. *Nat. Phys.* **18**, 112–117 (2022).
154. Chen, D. et al. Automating crystal-structure phase mapping by combining deep learning with constraint reasoning. *Nat. Mach. Intell.* **3**, 812–822 (2021).
155. Gómez-Bombarelli, R. et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* **4**, 268–276 (2018).
156. Anishchenko, I. et al. De novo protein design by deep network hallucination. *Nature* **600**, 547–552 (2021).
157. Fu, T. et al. Differentiable scaffolding tree for molecular optimization. In *International Conference on Learning Representations* (2021).
158. Sanchez-Lengeling, B. & Aspuru-Guzik, A. Inverse molecular design using machine learning: generative models for matter engineering. *Science* **361**, 360–365 (2018).
159. Huang, K. et al. Therapeutics Data Commons: machine learning datasets and tasks for drug discovery and development. In *NeurIPS Datasets and Benchmarks* (2021).
This study describes an initiative with open AI models, datasets and education programmes to facilitate advances in therapeutic science across all stages of drug discovery and development.
160. Dance, A. Lab hazard. *Nature* **458**, 664–665 (2009).
161. Segler, M. H. S., Preuss, M. & Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **555**, 604–610 (2018).
This paper describes an approach that combines deep neural networks with Monte Carlo tree search to plan chemical synthesis.
162. Gao, W., Raghavan, P. & Coley, C. W. Autonomous platforms for data-driven organic synthesis. *Nat. Commun.* **13**, 1075 (2022).
163. Kusne, A. G. et al. On-the-fly closed-loop materials discovery via Bayesian active learning. *Nat. Commun.* **11**, 5966 (2020).
164. Gormley,A. J. & Webb, M. A. Machine learning in combinatorial polymer chemistry. *Nat. Rev. Mater.* **6**, 642–644 (2021).
165. Ament, S. et al. Autonomous materials synthesis via hierarchical active learning of nonequilibrium phase diagrams. *Sci. Adv.* **7**, eabg4930 (2021).
166. Degrave, J. et al. Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature* **602**, 414–419 (2022).
This paper describes an approach for controlling tokamak plasmas, using a reinforcement-learning agent to command-control coils and satisfy physical and operational constraints.
167. Melnikov, A. A. et al. Active learning machine learns to create new quantum experiments. *Proc. Natl Acad. Sci. USA* **115**, 1221–1226 (2018).
168. Smith, J. S., Isayev, O. & Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chem. Sci.* **8**, 3192–3203 (2017).
169. Wang, D. et al. Efficient sampling of high-dimensional free energy landscapes using adaptive reinforced dynamics. *Nat. Comput. Sci.* **2**, 20–29 (2022).
This paper describes a neural network for reliable uncertainty estimations in molecular dynamics, enabling efficient sampling of high-dimensional free energy landscapes.
170. Wang, W. & Gómez-Bombarelli, R. Coarse-graining auto-encoders for molecular dynamics. *npj Comput. Mater.* **5**, 125 (2019).
171. Hermann, J., Schätzle, Z. & Noé, F. Deep-neural-network solution of the electronic Schrödinger equation. *Nat. Chem.* **12**, 891–897 (2020).
This paper describes a method to learn the wavefunction of quantum systems using deep neural networks in conjunction with variational quantum Monte Carlo.
172. Carleo, G. & Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **355**, 602–606 (2017).
173. Em Karniadakis, G. et al. Physics-informed machine learning. *Nat. Rev. Phys.* **3**, 422–440 (2021).
174. Li, Z. et al. Physics-informed neural operator for learning partial differential equations. Preprint at https://arxiv.org/abs/2111.03794 (2021).
175. Kochkov, D. et al. Machine learning–accelerated computational fluid dynamics. *Proc. Natl Acad. Sci. USA* **118**, e2101784118 (2021).
This paper describes an approach to accelerating computational fluid dynamics by training a neural network to interpolate from coarse to fine grids and generalize to varying forcing functions and Reynolds numbers.
176. Ji, W. et al. Stiff-PINN: physics-informed neural network for stiff chemical kinetics. *J. Phys. Chem. A* **125**, 8098–8106 (2021).
177. Smith, J. D., Azizzadenesheli, K. & Ross, Z. E. EikoNet: solving the Eikonal equation with deep neural networks. *IEEE Trans. Geosci. Remote Sens.* **59**, 10685–10696 (2020).
178. Waheed, U. B. et al. PINNeik: Eikonal solution using physics-informed neural networks. *Comput. Geosci.* **155**, 104833 (2021).
179. Chen, R. T. Q. et al. Neural ordinary differential equations. In *Advances in Neural Information Processing Systems* **31**, 6572–6583 (2018).
This paper established a connection between neural networks and differential equations by introducing the adjoint method to learn continuous-time dynamical systems from data, replacing backpropagation.
180. Raissi, M., Perdikaris, P. & Karniadakis, G. E. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.* **378**, 686–707 (2019).
This paper describes a deep-learning approach for solving forwards and inverse problems in nonlinear partial differential equations and can find solutions to differential equations from data.
181. Lu, L. et al. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nat. Mach. Intell.* **3**, 218–229 (2021).
182. Brandstetter, J., Worrall, D. & Welling, M. Message passing neural PDE solvers. In *International Conference on Learning Representations* (2022).
183. Noé, F. et al. Boltzmann generators: sampling equilibrium states of many-body systems with deep learning. *Science* **365**, eaaw1147 (2019).
This paper presents an efficient sampling algorithm using normalizing flows to simulate equilibrium states in many-body systems.
184. Rezende, D. & Mohamed, S. Variational inference with normalizing flows. In *International Conference on Machine Learning* **37**, 1530–1538, (2015).
185. Dinh, L., Sohl-Dickstein, J. & Bengio, S. Density estimation using real NVP. In *International Conference on Learning Representations* (2017).
186. Nicoli, K. A. et al. Estimation of thermodynamic observables in lattice field theories with deep generative models. *Phys. Rev. Lett.* **126**, 032001 (2021).
187. Kanwar, G. et al. Equivariant flow-based sampling for lattice gauge theory. *Phys. Rev. Lett.* **125**, 121601 (2020).
188. Gabrié, M., Rotskoff, G. M. & Vanden-Eijnden, E. Adaptive Monte Carlo augmented with normalizing flows. *Proc. Natl Acad. Sci. USA* **119**, e2109420119 (2022).
189. Jasra, A., Holmes, C. C. & Stephens, D. A. Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Stat. Sci.* **20**, 50–67 (2005).
190. Bengio, Y. et al. Better mixing via deep representations. In *International Conference on Machine Learning* 552–560 (2013).
191. Pompe, E., Holmes, C. & Łatuszyński, K. A framework for adaptive MCMC targeting multimodal distributions. *Ann. Stat.* **48**, 2930–2952 (2020).
192. Townshend, R. J. L. et al. ATOM3D: tasks on molecules in three dimensions. In *NeurIPS Datasets and Benchmarks* (2021).
193. Kearnes, S. M. et al. The open reaction database. *J. Am. Chem. Soc.* **143**, 18820–18826 (2021).
194. Chanussot, L. et al. Open Catalyst 2020 (OC20) dataset and community challenges. *ACS Catal.* **11**, 6059–6072 (2021).
195. Brown, N. et al. GuacaMol: benchmarking models for de novo molecular design. *J. Chem. Inf. Model.* **59**, 1096–1108 (2019).
196. Notin, P. et al. Tranception: protein fitness prediction with autoregressive transformers and inference-time retrieval. In *International Conference on Machine Learning* 16990–17017 (2022).
197. Mitchell, M. et al. Model cards for model reporting. In *Conference on Fairness, Accountability, and Transparency*220–229 (2019).
198. Gebru, T. et al. Datasheets for datasets. *Commun. ACM* **64**, 86–92 (2021).
199. Bai, X. et al. Advancing COVID-19 diagnosis with privacy-preserving collaboration in artificial intelligence. *Nat. Mach. Intell.* **3**, 1081–1089 (2021).
200. Warnat-Herresthal, S. et al. Swarm learning for decentralized and confidential clinical machine learning. *Nature* **594**, 265–270 (2021).
201. Hie, B., Cho, H. & Berger, B. Realizing private and practical pharmacological collaboration. *Science* **362**, 347–350 (2018).
202. Rohrbach, S. et al. Digitization and validation of a chemical synthesis literature database in the ChemPU. *Science* **377**, 172–180 (2022).
203. Gysi, D. M. et al. Network medicine framework for identifying drug-repurposing opportunities for COVID-19. *Proc. Natl Acad. Sci. USA* **118**, e2025581118 (2021).
204. King, R. D. et al. The automation of science. *Science* **324**, 85–89 (2009).
205. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
206. Doerr, S. et al. TorchMD: a deep learning framework for molecular simulations. *J. Chem. Theory Comput.* **17**, 2355–2363 (2021).
207. Schoenholz, S. S. & Cubuk, E. D. JAX MD: a framework for differentiable physics. In *Advances in Neural Information Processing* Systems **33**, 11428–11441 (2020).
208. Peters, J., Janzing, D. & Schölkopf, B. *Elements of Causal Inference: Foundations and Learning Algorithms* (MIT Press, 2017).
209. Bengio, Y. et al. A meta-transfer objective for learning to disentangle causal mechanisms. In *International Conference on Learning Representations* (2020).
210. Schölkopf, B. et al. Toward causal representation learning. *Proc. IEEE* **109**, 612–634 (2021).
211. Goyal, A. & Bengio, Y. Inductive biases for deep learning of higher-level cognition. *Proc. R. Soc. A* **478**, 20210068 (2022).
212. Deleu, T. et al. Bayesian structure learning with generative flow networks. In *Conference on Uncertainty in Artificial Intelligence* 518–528 (2022).

213. Geirhos, R. et al. Shortcut learning in deep neural networks. *Nat. Mach. Intell.* **2**, 665–673 (2020).
214. Koh, P. W. et al. WILDS: a benchmark of in-the-wild distribution shifts. In *International Conference on Machine Learning* 5637–5664 (2021).
215. Luo, Z. et al. Label efficient learning of transferable representations across domains and tasks. In *Advances in Neural Information Processing Systems* **30**, 165–177 (2017).
216. Mahmood, R. et al. How much more data do I need? estimating requirements for downstream tasks. In *IEEE Conference on Computer Vision and Pattern Recognition* 275–284 (2022).
217. Coley, C. W., Eyke, N. S. & Jensen, K. F. Autonomous discovery in the chemical sciences part II: outlook. *Angew. Chem. Int. Ed.* **59**, 23414–23436 (2020).
218. Gao, W. & Coley, C. W. The synthesizability of molecules proposed by generative models. *J. Chem. Inf. Model.* **60**, 5714–5723 (2020).
219. Kogler, R. et al. Jet substructure at the Large Hadron Collider. *Rev. Mod. Phys.* **91**, 045003 (2019).
220. Acosta, J. N. et al. Multimodal biomedical AI. *Nat. Med.* **28**, 1773–1784 (2022).
221. Alayrac, J.-B. et al. Flamingo: a visual language model for few-shot learning. In *Advances in Neural Information Processing Systems* **35**, 23716–23736 (2022).
222. Elmarakeby, H. A. et al. Biologically informed deep neural network for prostate cancer discovery. *Nature* **598**, 348–352 (2021).
223. Qin, Y. et al. A multi-scale map of cell structure fusing protein images and interactions. *Nature* **600**, 536–542 (2021).
224. Schaffer, L. V. & Ideker, T. Mapping the multiscale structure of biological systems. *Cell Systems* **12**, 622–635 (2021).
225. Stiglic, G. et al. Interpretability of machine learning-based prediction models in healthcare. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **10**, e1379 (2020).
226. Erion, G. et al. A cost-aware framework for the development of AI models for healthcare applications. *Nat. Biomed. Eng.* **6**, 1384–1398 (2022).
227. Lundberg, S. M. et al. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nat. Biomed. Eng.* **2**, 749–760 (2018).
228. Sanders, L. M. et al. Beyond low Earth orbit: biological research, artificial intelligence, and self-driving labs. Preprint at https://arxiv.org/abs/2112.12582 (2021).
229. Gagne, D. J. II et al. Interpretable deep learning for spatial analysis of severe hailstorms. *Mon. Weather Rev.* **147**, 2827–2845 (2019).
230. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **1**, 206–215 (2019).
231. Koh, P. W. & Liang, P. Understanding black-box predictions via influence functions. In *International Conference on Machine Learning* 1885–1894 (2017).
232. Mirzasoleiman, B., Bilmes, J. & Leskovec, J. Coresets for data-efficient training of machine learning models. In *International Conference on Machine Learning* 6950–6960 (2020).
233. Kim, B. et al. Interpretability beyond feature attribution: quantitative testing with concept activation vectors (TCAV). In *International Conference on Machine Learning* 2668–2677 (2018).
234. Silver, D. et al. Mastering the game of go without human knowledge. *Nature* **550**, 354–359 (2017).
235. Baum, Z. J. et al. Artificial intelligence in chemistry: current trends and future directions. *J. Chem. Inf. Model.* **61**, 3197–3212 (2021).
236. Finlayson, S. G. et al. Adversarial attacks on medical machine learning. *Science* **363**, 1287–1289 (2019).
237. Urbina, F. et al. Dual use of artificial-intelligence-powered drug discovery. *Nat. Mach. Intell.* **4**, 189–191 (2022).
238. Norgeot, B. et al. Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist. *Nat. Med.* **26**, 1320–1324 (2020).

Acknowledgements M.Z. gratefully acknowledges the support of the National Institutes of Health under R01HD108794, U.S. Air Force under FA8702-15-D-0001, awards from Harvard Data Science Initiative, Amazon Faculty Research, Google Research Scholar Program, Bayer Early Excellence in Science, AstraZeneca Research, Roche Alliance with Distinguished Scientists, and Kempner Institute for the Study of Natural and Artificial Intelligence. C.P.G. and Y.D. acknowledge the support from the U.S. Air Force Office of Scientific Research under Multidisciplinary University Research Initiatives Program (MURI) FA9550-18-1-0136, Defense University Research Instrumentation Program (DURIP) FA9550-21-1-0316, and awards from Scientific Autonomous Reasoning Agent (SARA), and AI for Discovery Assistant (AIDA). Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funders. We thank D. Hassabis, A. Davies, S. Mohamed, Z. Li, K. Ma, Z. Qiao, E. Weinstein, A. V. Weller, Y. Zhong and A. M. Brandt for discussions on the paper.

Author contributions All authors contributed to the design and writing of the paper, helped shape the research, provided critical feedback, and commented on the paper and its revisions. H.W., T.F., Y.D. and M.Z conceived the study and were responsible for overall direction and planning. W.G., K.H. and Z.L. contributed equally to this work (equal second authorship) and are listed alphabetically.

Competing interests The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to Marinka Zitnik.
Peer review information *Nature* thanks Brian Gallagher and Benjamin Nachman for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.