Nanoscale

REVIEW

View Article Online
View Journal

Cite this: DOI: 10.1039/d5nr02043c

Data driven approaches in nanophotonics: a review of AI-enabled metadevices

Huanshu Zhang, Lei Kang, * Sawyer D. Campbell, Jacob T. Young and Douglas H. Werner *

Data-driven approaches have revolutionized the design and optimization of photonic metadevices by harnessing advanced artificial intelligence methodologies. This review takes a model-centric perspective that synthesizes emerging design strategies and delineates how traditional trial-and-error and computationally intensive electromagnetic simulations are being supplanted by deep learning frameworks that efficiently navigate expansive design spaces. We discuss artificial intelligence implementation in several metamaterial design aspects from high-degree-of-freedom design to large language model-assisted design. By addressing challenges such as transformer model implementation, fabrication limitations, and intricate mutual coupling effects, these AI-enabled strategies not only streamline the forward modeling process but also offer robust pathways for the realization of multifunctional and fabrication-friendly nanophotonic devices. This review further highlights emerging opportunities and persistent challenges, setting the stage for next-generation strategies in nanophotonic engineering.

Received 15th May 2025,
Accepted 26th September 2025

DOI: 10.1039/d5nr02043c

rsc.li/nanoscale

Introduction

The emergence of artificial structures not only unveils the complexity of light–matter interaction, which generally occurs within atoms or molecules, but also provides an unprecedented opportunity for control of light *via* engineered devices. Compared with photonic crystals which rely on collective response arising from periodic perturbations, metamaterials have demonstrated powerful and versatile manipulation of electromagnetic (EM) waves *via* structural engineering of their subwavelength building blocks, *i.e.*, meta-atoms.[1] For instance, metamaterials exhibiting properties that are not usually seen in naturally occurring materials, such as optical magnetism,[2] negative effective refractive index,[3] and strong chirality[4] have been reported. Importantly, those exotic properties primarily arise from a meta-atoms' architecture rather than the intrinsic properties of the base materials. These characteristics make metamaterials an excellent candidate for applications in devices that require precise control over the intrinsic properties (magnitude, phase, polarization, *etc.*) of light. Extending these principles to 2D has led to the development of metasurfaces which gain their properties from single-layer or few-layer planar artificial structures. In contrast to metamaterials, metasurfaces provide a more compact and fabrication-friendly approach to light control,[5] facilitating applications such as wavefront shaping,[6] beam steering,[7] holography,[8] optical computing,[9] *etc.* The development of metadevices as an extension of the metamaterial and metasurface paradigm paves the way toward the next generation of photonic technologies.

Though metamaterials and metasurfaces have manifested an impressive capability to manipulate light, the design of metadevices, especially those for sophisticated and/or multiple functionalities can be an extremely challenging engineering task. Metamaterial design has traditionally relied on iterative numerical simulations that solve Maxwell's equations to determine the optical/electromagnetic response of its constituent meta-atoms. Common computational tools include the Finite-Difference Time-Domain (FDTD) method,[10] and Finite-Element Method (FEM),[11] as they discretize EM fields in space and/or time to analyze how nanostructures interact with EM waves. However, given the fact that meta-atoms generally include deep-subwavelength features, accurate metamaterial simulations based on these methods can be computationally expensive,[12] especially when considering multiscale devices. On the other hand, optimization methods such as genetic algorithms (GA) or other evolutionary optimization techniques are commonly employed to design metamaterials structures.[13] The optimization iteratively refines designs by evaluating optical properties of the structure using numerical EM solvers (commercial or customized), updating the design based on an objective function, and repeating the process until some stopping criteria is met.[14] Despite previous efforts made to improve the efficiency of both optimization methods and EM

Department of Electrical Engineering, The Pennsylvania State University, University Park, PA 16801, USA. E-mail: lzk12@psu.edu, dhw@psu.edu

solvers, these approaches still impose a significant computational burden in terms of time and resources. Consequently, evaluating and optimizing subwavelength architecture remains a great challenge, necessitating more advanced and efficient strategies.

Machine learning (ML) and deep learning (DL), subsets of artificial intelligence (AI),[15] offer data-driven approaches to identifying complex structure/response correlations in metamaterial design due to their ability to replicate hyper-dimensional non-linear relationships and, once trained, near instantaneous evaluation. ML algorithms include supervised, unsupervised, and reinforcement learning, with supervised learning being the most utilized due to its structured data-label relationships.[16] Fully connected (FC) neural networks, such as multilayer perceptrons (MLPs), have been used to map design parameters to optical properties.[17–20] Recurrent neural networks (RNNs), including long short-term memory (LSTM)[21] networks, excel in sequential data, making them suitable for applications involving sequence processing such as in the case of spectra.[22,23] Convolutional neural networks (CNNs), designed for image processing, effectively analyze spatial patterns within metasurfaces.[24,25] The transformer architecture, first introduced by Vaswani *et al.* in 2017,[26] uses self-attention mechanisms to capture complex data dependencies, enabling parallelized training and specializing in sequence modelling, forming the basis for large language models (LLMs) like ChatGPT,[27] has also been applied in designing metamaterials.[28]

DL has rapidly advanced metamaterial modelling and optimization.[29,30] The journey began in 2017 when Malkiel *et al.* developed a bidirectional MLP-based DL model for designing H-shaped plasmonic nanostructures.[17] Ma *et al.* (2018) further advanced the method by developing two bidirectional neural networks with partial stacking for both forward and inverse modelling of reflection and CD spectra,[19] while Peurifoy *et al.* (2018) applied MLPs to optimize multilayer nanoparticles.[20] Asano *et al.* (2018) employed CNNs to enhance the *Q*-factor of photonic crystals.[31] Sajedian *et al.* (2019) integrated CNNs and RNNs to predict absorption spectra of random plasmonic nanostructures.[32] Chen *et al.* (2023) recently harnessed a transformer-based model for designing broadband solar metamaterial absorbers.[28]

A few recent review papers on AI-assisted metamaterial design have charted diverse paths toward solving forward and inverse design challenges. For instance, Masson *et al.* (2023) have discussed the use of ML for nanoplasmonics, revealing how advanced algorithms uncover the complex structure–property relationships that underpin high-performance device engineering.[33] Chen *et al.* (2022) have bridged the fields of AI and meta-optics by detailing how AI accelerates both design and functional realization of flat optical devices.[34] Furthermore, Ueno *et al.* (2024) have offered the perspective that zeroes in on AI-enabled design-for-manufacturing and computational post-processing, help mitigate the simulation–fabrication gaps in metasurfaces.[29] Other review articles[12,30,35–47] have been dedicated to different aspects such as free-form optimization,[48] light–matter interactions,[49] physics-informed neural networks,[50] and intelligent inverse design for phononic metamaterials.[37] In contrast, from a model-centric perspective, this review will primarily focus on emerging design strategies. In particular, we aim to discuss high-degree-of-freedom (DoF) metamaterial design, the use of transformers and attention mechanisms, the prediction of mutual coupling effects between meta-atoms, strategies for designing robust and fabrication-friendly metamaterials, and recent advances as well as future prospects. We also summarize the computational costs in practice for those reported methods. Across the studies we survey, the dominant expense is usually dataset generation (*e.g.,* full-wave simulations), while model training is a secondary, one-off cost and inference is typically near-interactive on commodity GPUs. For inverse design, simpler models such as one-shot/tandem/autoencoder-based approaches tend to yield faster per-candidate sampling than sequential samplers (*e.g.,* diffusion), which trade speed for stability/diversity. Consequently, hardware requirements scale most strongly with training-set size, fidelity and the inverse-design strategy, rather than with the specific deep-learning library. In practice, a single consumer GPU is usually sufficient for training and inference, whereas multi-GPU or cluster access mainly benefits large-scale dataset generation or LLMs training. To enable apples-to-apples comparisons, we encourage future reports to specify (i) dataset-generation setup, (ii) training wall-clock and memory footprint, and (iii) per-sample inference cost.

AI-assisted high-DoF metamaterial design

Image-based methods

Most ML-based approaches for AI-assisted forward design of high-DoF metamaterials function as black-box models, where the metamaterial structure (geometries and material properties) serves as input, and the optical response (for instance, transmission and reflection coefficients) is predicted as output. In this context, 'high-DoF' refers to a hyper-dimensional design space characterized by independently tunable parameters (*e.g.*, 10+), making traditional iterative method unfeasible. To encode the metamaterial unit cell into a format suitable for ML models, two primary digitalization methods have been adopted.[51] The first method involves pixelating the planar metamaterial and treating the resulting representation as an image, often in a binary format. Computer vision techniques, particularly CNNs, are then applied to optimize the pixel distribution. This approach allows for high-DoF designs, as each pixel can be independently adjusted to form intricate structures. Moreover, the pixel-based representation aligns with human intuition, as the unit cell's geometry is directly visualized, facilitating the interpretation of optimization outcomes. Research efforts have explored and refined this approach to enhance metamaterial design efficiency and performance.[35,41,52–64]

One early demonstration of image-driven high-DoF metamaterial modelling was provided by An *et al.* (2020), who developed a CNN to predict wideband amplitude and phase responses of quasi-freeform dielectric metasurfaces[25] (Fig. 1(a)). They leveraged the CNN model's ability to handle structures across varying lattice constants, material indices, and thicknesses. They achieved an average prediction standard deviation of 0.005 (amplitude) and 0.78 degrees (phase) at each single frequency point after training on more than 100 000 simulation data sets. This study underscored the viability of image-based approaches to accelerate metamaterial designs, opening the door to fast performance evaluation for high-DoF structures. However, generating such an enormous training data set can be extremely time-consuming, raising concerns about the practical efficiency. One might question whether conventional simulation-optimization methods could achieve comparable performance with fewer simulations, calling into question the trade-off between model accuracy and

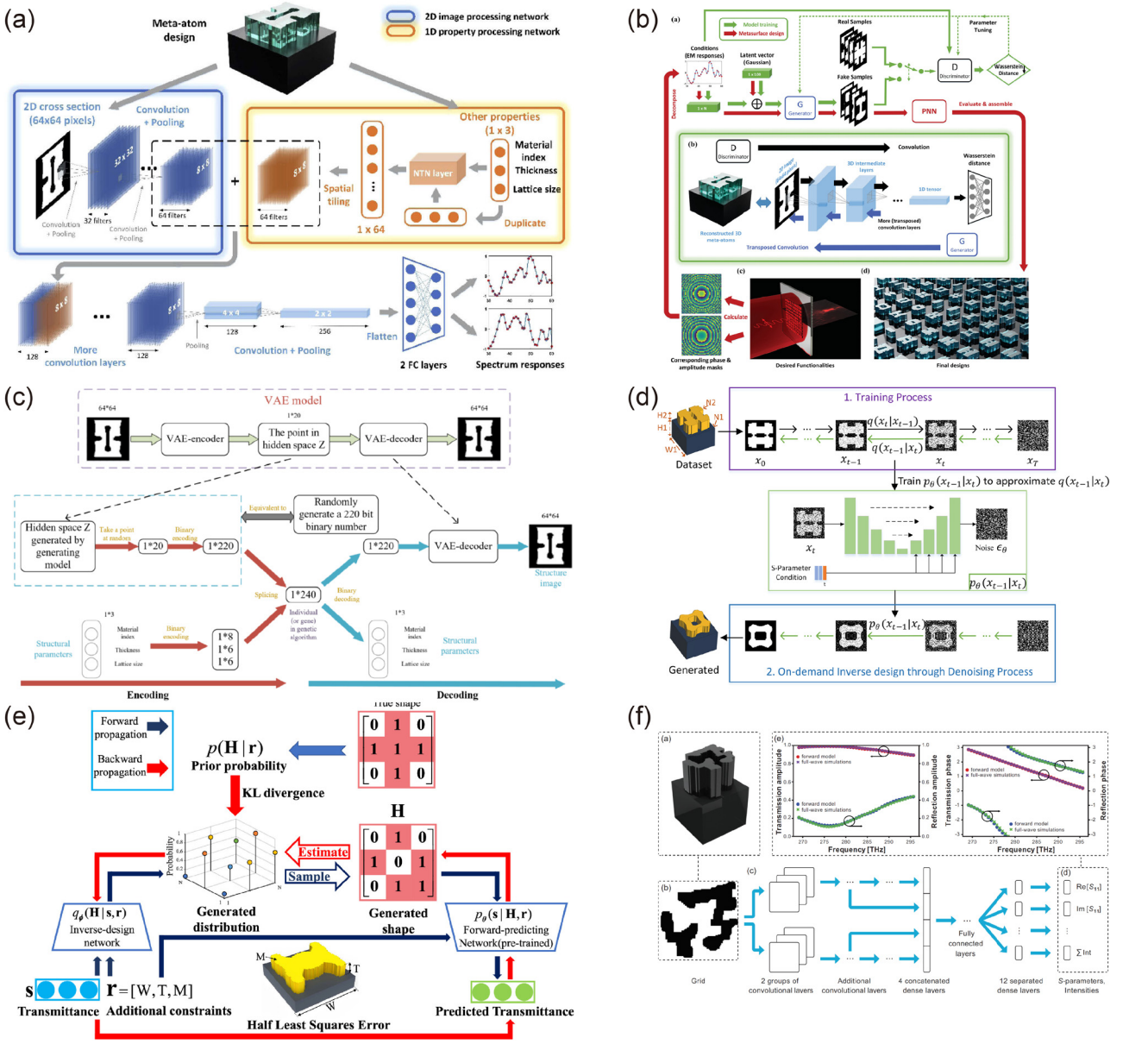

Fig. 1 Image-based method for AI-assisted design of metasurfaces. (a) CNN network for design of high-DoF quasi-freeform dielectric metasurfaces. Reprinted with permission from ref. 25. Copyright 2020, Optical Society of America. (b) GAN model for metasurface inverse design. Reprinted with permission from ref. 65. Copyright 2021, Wiley-VCH. (c) VAE and GA for metasurface inverse design. Reprinted with permission from ref. 66. Copyright 2022, Optical Society of America. (d) The first diffusion probabilistic model for inverse design of meta-atoms. Reprinted with permission from ref. 67. The article is licensed under a Creative Commons Attribution 4.0 International License https://creativecommons.org/licenses/by/4.0/. (e) The first probabilistic generative model in a tandem architecture (TGN) for the design of meta-atoms. Reprinted with permission from ref. 68. Copyright 2025 American Chemical Society. (f) 100 × 100 binary images were used for freeform metasurfaces forward and inverse design. Reprinted with permission from ref. 69. Copyright 2022 American Physical Society https://doi.org/10.1103/PhysRevB.106.085408.

data set size. Notably, while training required ~48 hours on two NVIDIA 1080 Ti GPUs, curating the >100k-sample corpus took ~8 days across six servers (~48 server-days), making data generation, not training, the dominant cost. Surrogate modelling only pays off when the model is reused extensively across many design queries; for one-off or small-batch tasks, the upfront data cost can outweigh the inference-time speedup.

This work was soon extended to exploit other network types. In 2021, An *et al.* introduced a Generative Adversarial Network (GAN) for the inverse design of quasi-freeform structures[65] (Fig. 1(b)), using Wasserstein GAN (WGAN) that learns to generate free-form dielectric meta-atom patterns conditioned on desired amplitude and phase responses. They produced 100 qualified designs in 32 seconds under a tight threshold of ±0.1 amplitude error and ±10° phase error in dual-target cases. For the field of AI-assisted metamaterials design, this work marked a turning point by proving that GAN-based models can directly synthesize meta-atom layouts targeting multifunctional properties in one step. In 2022, Yu *et al.* further improved An's work by combining a Variational AutoEncoder (VAE) and GA for inverse design[66] (Fig. 1(c)). Compared to GANs, the VAE compresses the large design space into a latent manifold of feasible structures, which acted as a search domain for a GA. By looping the GA over the VAE's learned manifold of solutions, the algorithm could escape local optima and eventually converge to a meta-atom configuration that met the target spectral requirements. The training process lasts about 6 h on two NVIDIA GeForce GTX 3080. In 2023, diffusion models entered the field: Zhang *et al.* proposed a diffusion probabilistic model for generating high-DOF meta-atom images conditioned on desired wideband *S*-parameter spectra[67] (Fig. 1(d)). Starting from random noise, the diffusion model iteratively refines the pixelated meta-atom design such that its simulated electromagnetic response converges to the specified target spectrum. This diffusion-driven strategy inherently avoids the training instabilities of GANs by eschewing adversarial objectives altogether. However, diffusion models require running additional sequential denoising steps to generate each design, which can significantly increase the computational time for inference. As a result, on-demand design generation is generally slower (0.43 s per design as reported) compared to direct one-shot mapping techniques (such as those based on VAE or tandem networks, about 1 ms per design on common commercial GPUs), as each solution must be iteratively computed, suggesting a trade-off between the speed for stability improvement and the accuracy in design outcomes. In 2025, Yang *et al.* introduced a probabilistic generative model in a tandem architecture, *i.e.* the Tandem Generative Network (TGN), for design of meta-atoms[68] (Fig. 1(e)). The TGN architecture couples a forward neural network (to capture physics of a meta-atom's response) with a generative network that samples new structure images from a learned probability space. This tandem setup addresses two key issues: the difficulty of handling one-to-many mappings in inverse problems (*e.g.,* multiple structures yielding similar spectra) and the slow generation speed typical of vanilla diffusion models. Claiming up to 38% lower mean absolute error (MAE) and nearly 3000× faster generation (generated 10 000 atoms in 3.73 s) than the diffusion model,[67] TGN represents a further step in improving both speed and precision in high-DoF metasurface design.

Beyond quasi-freeform-related dielectric metasurfaces, several studies have applied image-processing networks to optimize other large, pixelated metasurfaces. For instance, Gahlmann and Tassin (2022) first trained a CNN to emulate the forward mapping from a 100 × 100 binary meta-atom image to its full spectra (*S*-parameters), then embedded this CNN into a conditional GAN (CGAN), which will propose new meta-atom images given target spectra[69] (Fig. 1(f)). Similarly, Li *et al.* (2022) proposed a CNN to predict the circular dichroism (CD) response of chiral metasurfaces with nanohole arrays (represented as 80 × 80 binary images).[24] In another approach, Tanriover *et al.* (2022) introduced an AutoEncoder (AE) model for designing free-form 100 × 100 binary meta-atom images.[70] Collectively, these studies underscore the potential of DL to efficiently handle both forward and inverse design challenges while respecting fabrication constraints.

The rapid progress in image-based methods for metamaterial design points toward a future where scalability, interpretability, fabrication integration, and data efficiency become the focal points of research. On the scalability front, next-generation models will need to handle volumetric metamaterials as input "images", which will enable the co-design of large-scale devices with many interacting elements without sacrificing resolution. Equally important is enhancing the interpretability of these models. As AI-designed metamaterials begin being put into practical use, designers will demand insights into how network features help to explain physical phenomena. Another critical direction is the seamless integration of manufacturing constraints and feedback into the design loop. Future AI models may incorporate differentiable fabrication-process simulators that ensure generated designs are not only nominally optimal but also robust against fabrication imperfections and material tolerances, which could involve training networks on experimental data. Additionally, emphasis on data efficiency will grow: instead of relying on tens of thousands of simulated examples, researchers are exploring physics-informed neural networks (PINNs), transfer learning, and active learning to make the most of limited data. We will cover the recent advances on these topics in our following discussions.

Parameter-based methods

While image-based methods have advanced the design of high-DoF planar metamaterials, their reliance on pixel discretization limits their implementation in *true* 3D meta-atoms with structural variations in the light propagation direction. To work around these constraints, parameter-based methods have been introduced as an alternative framework. In this approach, each meta-atom is represented by a vector that encodes its geometry and material properties, enabling a description of 3D subwavelength architectures. By shifting from a discrete pixel representation to a parameter space, these methods facilitate designs that

capture intricate spatial details absent in image-based models. Sequence-to-sequence models, including FC layers and RNNs like Gated Recurrent Units (GRUs)[71] and LSTMs,[21] are often used to process these parameter vectors. The more recently emerged transformer architectures with self-attention mechanisms offered new possibilities for efficiently mapping complex parameter spaces to optical responses.[28]

In their pioneering work published in 2017, Malkiel *et al.* introduced a DL algorithm for plasmonic nanostructure design using an 8-parameter model to generate H-shaped planar structures within a design space comprising approximately 2.33×10^8 possible configurations, addressing the long-standing challenge of time-consuming numerical simulation[17,72] (Fig. 2(a)). By training a bidirectional Deep Neural Network (DNN) composed of multiple FC layers on over 15 000 simulation instances, the authors achieved a transmission spectra prediction mean squared error (MSE) of 0.16. This study was significant as it was one of the first to solve a nontrivial metasurface design problem with DL, offering orders-of-magnitude speedups over iterative solvers. In 2018, Peurifoy *et al.* showed that a DNN can approximate the forward light-scattering behaviours of multilayered nanoparticle structures with high accuracy using a relatively small training set[20] (Fig. 2(b)). Once trained, their network reached a mean relative error (MRE) of 1.5%. In a later study (2019), Nadell *et al.* applied DL to model and design all-dielectric metasurfaces, which involve multiple resonant modes and near-field coupling between elements[73] (Fig. 2(c)). Their method, which incorporated not only the raw parameters but also their ratios as inputs, achieved a transmittance prediction MSE of 1.16×10^{-3} after training on 18 000 simulation samples (about 0.1 ms per prediction on Tesla Quadro M6000). This study validated that DNNs can handle relatively large, complex unit cells. In 2021, Xu *et al.* combined NNs with transfer learning and GA to design phase-modulating metasurfaces[74] (Fig. 2(d)). In their approach, a forward spectrum-prediction network was first trained on a base task (a rectangular meta-atom) and then fine-tuned (transferred) to a new task (elliptical meta-atom) using far fewer samples. The enhanced accuracy from transfer learning allowed the network to serve as a high-fidelity surrogate model for a genetic algorithm. The significance of this work lies in its hybrid strategy: by reducing the NN's data requirements and integrating it with GA, it demonstrated a practical route to designing large-area functional metasurfaces more quickly. In 2022, Liao *et al.* further extended AI-based design to 3D chiral plasmonic metasurfaces[75] (Fig. 2(e)). In practice, separate models were trained for a given handedness of a chiral structure; then the knowledge learned was partially transferred to a new model for the opposite handedness, greatly accelerating convergence while requiring little additional data. This study demonstrated that even highly intricate design spaces (like 3D chiral nanoresonators with polarization-dependent responses) can be handled by DL when augmented with physics-informed training (transfer learning between related design tasks) and feature-extraction methods. Looking to the current state of the art, researchers are pushing parameter-based methods to handle higher-dimensional design spaces and more complex unit cell architectures. In 2025, Zhang *et al.* introduced a fixed-attention LSTM-based approach for both forward and inverse design of true 3D plasmonic metamaterials, defined by 12 parameters (representing two gold nanorods embedded in a dielectric substrate), thereby exploring a design space of approximately 3.09×10^{19} possible configurations (Fig. 2(f)).[76] They treated the ordered list of design parameters as a temporal sequence and learned to "focus" to the most influential parameters during training. This attention-enhanced LSTM achieved ~48% lower MSE on the metasurface's transmission compared to a standard LSTM without attention, with about 3 ms per prediction on one NVIDIA GeForce RTX 2080 Ti. This work addresses the "curse of dimensionality" in metamaterial design by intelligently structuring the network to handle many design variables (and their interdependencies), it opens the door for AI-assisted optimization of high-DoF metamaterials that were previously intractable.

In contrast to the studies mentioned above in this section, several recent studies have employed models based on a limited number of design parameters. For instance, Ma *et al.* (2018) proposed a DL model for the design of stacked, twisted gold split ring resonators (SRRs) with dielectric spacers[19] (Fig. 2(g)). After training on 25 000 samples, they achieved an MSE of 1.6×10^{-4} for reflection amplitude predictions. Liu *et al.* (2018) directly confronted the one-to-many mapping issue inherent in photonic inverse design, where multiple distinct structures can exhibit the same spectrum[77] (Fig. 2(h)). They introduced a tandem neural network training approach in which an inverse-design network is cascaded with a pre-trained forward network (kept fixed) during training. Instead of learning an arbitrary mapping from spectrum to a particular geometry, the inverse model learns to produce any geometry that yields the desired spectrum by minimizing the error between the forward-predicted spectrum of its output design and the target spectrum. By effectively bypassing the need for one-to-one training pairs, this strategy enabled stable training of inverse models on datasets containing non-unique (*i.e.*, degenerate) solutions, paving the way for reliable DL-based design of more complex photonic structures without being hindered by mode degeneracies or ambiguous mappings. Mall *et al.* (2020) introduced a bidirectional AE (biAE) for plasmonic metasurfaces (with 4 structural parameters) that generated 10^5 possible design configurations[78] (Fig. 2(i)), reaching an MAE of 1.43% on validation cases after training on 1200 full-wave simulation examples. Hou *et al.* (2020) applied a tandem network for metamaterial absorbers defined by 6 parameters, achieving a test-set MSE of 2.95×10^{-4} after training on 20 000 samples (predicted in milliseconds per design on one NVIDIA GTX1060)[79] (Fig. 2(j)). Later studies by Han *et al.* (2023) and Luo *et al.* (2024) employed similar architectures (using 4 parameters) for designing chiral metastructures, with error metrics on the order of 10^{-4} (Fig. 2(k and l)).[80,81]

In Table 1, we summarize the recent research on the degrees of freedom and the corresponding network architec-

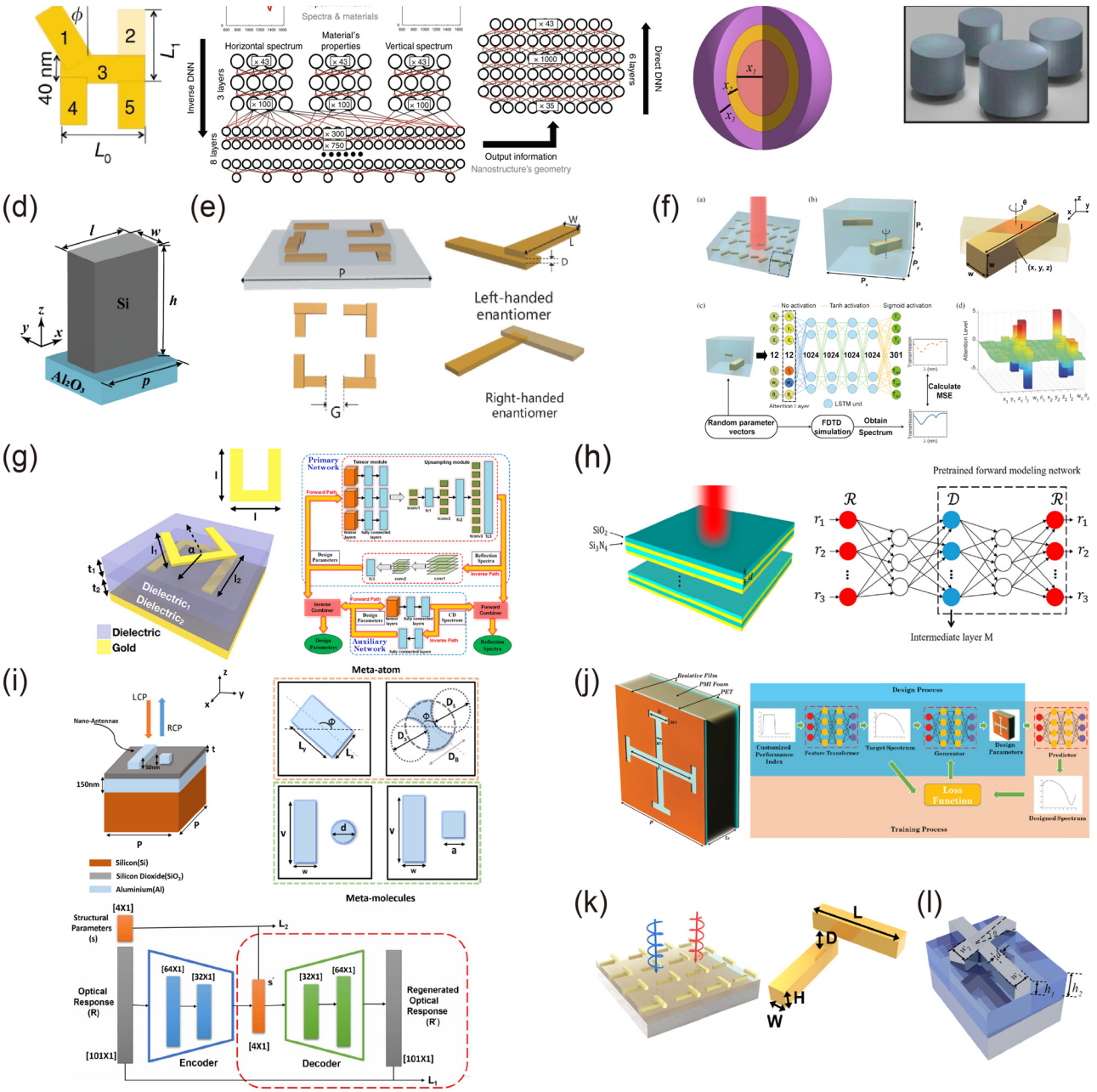

Fig. 2 Parameter-based method for AI-assisted design of metasurfaces. (a) An H-shaped planar structure design using DNN. Reproduced with permission from ref. 72. The article is licensed under a Creative Commons Attribution 4.0 International License https://creativecommons.org/licenses/by/4.0/. (b) Multilayered nanoparticle structures. Reproduced from ref. 20 © 2018 The Authors, with permission under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0) https://creativecommons.org/licenses/by-nc/4.0/. (c) An all-dielectric metasurface with multiple resonant modes and near-field coupling between elements. Reprinted with permission from ref. 73. Copyright 2019, Optical Society of America. (d) A rectangular-shaped phase-modulating meta-structure. Reprinted with permission from ref. 74. Copyright 2021, Optical Society of America. (e) A 3D Born–Kuhn type chiral metasurface. Reprinted with permission from ref. 75. Copyright 2022, Optical Society of America. (f) AI-assisted true 3D plasmonic high-DoF metamaterials design. Reproduced with permission from ref. 76. Copyright 2025 Optical Society of America. (g) A stacked, twisted gold split ring resonator with dielectric spacers. Reproduced with permission from ref. 19. Copyright 2018 American Chemical Society. (h) Tandem networks were used to design thin-film metasurfaces. Reproduced with permission from ref. 77. Copyright 2018 American Chemical Society. (i) Metal-dielectric–metal periodic gap-plasmon based half-wave plate metasurface design based on biAE. Reproduced with permission from ref. 78. Copyright 2020 Institute of Physics. (j) Metamaterial absorber design based on tandem networks. Reproduced with permission from ref. 79. The article is licensed under a Creative Commons License https://creativecommons.org/licenses/by/4.0/. (k) A chiral plasmonic Born–Kuhn metamaterial design based on multi-task learning. Reproduced with permission from ref. 80. Copyright 2023 American Chemical Society. (l) A dagger-shaped Ag array and an Ag mirror separated by a dielectric spacer. Reprinted with permission from ref. 81. Copyright 2024, Optical Society of America.

Table 1 Comparison of forward-design accuracy in parameter-based techniques

Works	Dimensions	Number of parameters	Design space	Training set	Network type	Error metrics
Malkiel *et al.* (2017)[17]	2D	8	2.33×10^{8} [a]	15 000	FC layers	MSE: 0.16
Peurifoy *et al.* (2018)[20]	1D	8	7.98×10^{12} [a]	50 000	FC layers	MRE: 1.5%
Ma *et al.* (2018)[19]	3D	5	Not mentioned	25 000	CNN	MSE: 1.6×10^{-4}
Liu *et al.* (2018)[77]	1D	16	6.57×10^{34} [a]	Not mentioned	FC layers	Error: 0.19
Nadell *et al.* (2019)[73]	2.5D	8	8.16×10^{8}	18 000	FC layers	MSE: 1.16×10^{-3}
An *et al.* (2019)[82]	2.5D	4	1.91×10^{11} [a]	35 000	FC layers	MSE: 3.5×10^{-4}
Gao *et al.* (2019)[83]	2.5D	4	8.70×10^{8}	3900	FC layers	MSE: 1.03×10^{-5}
Lin *et al.* (2019)[84]	2.5D	4	4.12×10^{9} [a]	25 900	FC layers	MSE: 1.04×10^{-3}
Li *et al.* (2019)[85]	2.5D	3	2.48×10^{7} [a]	2254	FC layers	MSE: 3.86×10^{-5}
Sajedian *et al.* (2019)[86]	2.5D	8	5.72×10^{9}	Not mentioned	FC layers	Not applied
Hou *et al.* (2020)[79]	2.5D	6	Not mentioned	20 000	FC layers	MSE: 2.95×10^{-4}
Unni *et al.* (2020)[87]	1D	11	1.79×10^{26} [a]	100 800	FC layers	Negative log-likelihood: −4.5
Mall *et al.* (2020)[78]	2D	4	1.00×10^{5}	1200	biAE	MAE: 1.43%
Tanriover *et al.* (2020)[88]	2.5D	3	1.41×10^{8} [a]	3157	FC layers	MSE: 2.1×10^{-3}
Qiu *et al.* (2020)[89]	1D	8	2.03×10^{7} [a]	80 000	FC layers	MSE: 5×10^{-2}
Xu *et al.* (2020)[90]	2D	3	Not mentioned	25 000	FC layers	MSE: 5×10^{-3}
Unni *et al.* (2021)[91]	1D	20	4.30×10^{39} [a]	579 600	FC layers	RMSE: 2×10^{-2}
Xu *et al.* (2021)[74]	2.5D	4	9.29×10^{8} [a]	27 000	FC layers	MSE: 7.7×10^{-4}
Lininger *et al.* (2021)[92]	1D	5	1.00×10^{12}	200 000	CNN	RMSE: 2×10^{-2}
Zandehshahvar *et al.* (2021)[93]	1D	8	7.98×10^{12} [a]	40 000	AE	MSE: 2.2×10^{-6}
Huang *et al.* (2021)[94]	2D	5	3.03×10^{14} [a]	12 040	FC layers	MSE: 1.4×10^{-2}
Sun *et al.* (2021)[95]	2D	4	5.94×10^{7} [a]	2.16×10^{7}	K-nearest neighbor (KNN)	MSE: 3.46×10^{-6}
Tanriover *et al.* (2021)[96]	2.5D	4	1.61×10^{11} [a]	6318	Complex valued FC layers	MSE: 1.2×10^{-4}
Deng *et al.* (2021)[97]	2.5D	14	1.04×10^{12}	24 000	CNN	MSE: 1.2×10^{-3}
Xu *et al.* (2021)[98]	2D	4	8.55×10^{7} [a]	71 808	FC layers	Accuracy: 96.49%
Noureen *et al.* (2022)[99]	2.5D	7	7.39×10^{12} [a]	Not mentioned	FC layers	MSE: 1.8×10^{-3}
Liao *et al.* (2022)[75]	2.5D	5	4.25×10^{7} [a]	23 000	FC layers	MSE: 1.6×10^{-4}
Gao *et al.* (2022)[100]	2D	4	2.08×10^{12} [a]	10 350	Modified FC layers	MSE: 1.47×10^{-4}
Shen *et al.* (2022)[101]	2D	5	1.38×10^{10} [a]	8400	FC layers	MSE: 1.29×10^{-4}
Deng *et al.* (2022)[22]	2.5D	4	1.35×10^{14} [a]	8000	LSTM	MAE: 8×10^{-2}
Lin *et al.* (2022)[102]	2D	16	4.30×10^{9}	55 000	FC layers	MSE: 3.24
Li *et al.* (2022)[103]	2D	5	3.12×10^{11} [a]	10 000	FC layers	MSE: less than 1×10^{-3}
Knightley *et al.* (2022)[104]	1D	13	2.92×10^{22} [a]	40 000	FC layers	MSE: 5×10^{-4}
Chen *et al.* (2022)[105]	2.5D	6	1.97×10^{14} [a]	4812	FC layers	MSE: 2.66×10^{-3}
Qiu *et al.* (2023)[106]	2D	4	3.51×10^{6} [a]	750	FC layers	MAE: 3×10^{-2}
Liu *et al.* (2023)[107]	2D	4	4.32×10^{9} [a]	80 000	FC layers	MSE: 8.7×10^{-3}
Jiang *et al.* (2023)[108]	2D	10	Not mentioned	528 000	FC layers with residual block	MSE: 1.3×10^{-5}
Yu *et al.* (2023)[109]	2.5D	7	2.18×10^{7}	8000	FC layers	MSE: 1.6×10^{-4}
Han *et al.* (2023)[80]	3D	4	Not mentioned	3075	FC layers	RMSE: 3.57×10^{-5}
Jahan *et al.* (2024)[110]	2.5D	5	3.38×10^{7}	5324	FC layers	MSE: 2.44×10^{-4}
Luo *et al.* (2024)[81]	2D	4	Not mentioned	7200	FC layers	MSE: 8.5×10^{-3}
Chen *et al.* (2024)[111]	2.5D	4	2.83×10^{6}	320	CNN with LSTM	MSE: 1.2×10^{-2}
Wang *et al.* (2024)[112]	2D	5	7.57×10^{8} [a]	18 144	FC layers	MSE: 2.8×10^{-4}
Zhu *et al.* (2024)[113]	2D	9	3.47×10^{13} [a]	4736	FC layers	MSE: 6×10^{-4}
Fan *et al.* (2024)[114]	2.5D	5	Not mentioned	2400	FC layers	MSE: 7.4×10^{-3}
Liu *et al.* (2025)[115]	2.5D	6	8.90×10^{6} [a]	12 988	FC layers	MSE: 3×10^{-5}
Yu *et al.* (2025)[116]	2.5D	5	Not mentioned	2400	FC layers	MSE: 1×10^{-3}
Chen *et al.* (2025)[117]	2.5D	16	1.00×10^{43}	Not mentioned	FC layers with Bayesian linear regression	Not mentioned
Zhang *et al.* (2025)[76]	3D	12	3.09×10^{19}	6393	LSTM with fixed-attention	MSE: 2.17×10^{-3}

[a] Indicates estimated based on parameter ranges provided in the paper.

tures. For the "Dimensions" column, the categories are defined as follows: (i) 1D: meta-atom's parameters vary along a single axis; (ii) 2D: meta-atoms can be represented as binary images. 2.5D: meta-atoms can be described using a binary image combined with one additional parameter representing thickness information; (iii) 3D: true 3D meta-atoms with structural variations in the vertical direction. Although studies on 1D structures can efficiently explore a vast design space, many such systems can be rapidly solved using conventional optimization methods, suggesting the unnecessity of using AI-assisted approaches. Parameter-based AI design methods for metamaterials will evolve in several directions. First, as unit cells become more complex, future models must handle potentially dozens of design variables without compromising accuracy or requiring extensive training data. This evolution may involve network architectures such as attention mechanisms, physics-informed neural networks, and modular networks, along with dimensionality reduction techniques that isolate

key design features. Transformer-based models and attention mechanisms offer a promising approach in this context. Also, multimodal learning that integrates electromagnetic simulations, experimental spectra, and fabrication constraints within a single model may further improve design efficiency. These advancements are expected to support the development of metamaterial systems that address practical design challenges and span high-dimensional design spaces.

Transformers and attentions applied in metamaterial design

Transformers are a modern DL architecture that relies on an attention mechanism to process information, rather than the convolution or recurrence used in traditional models (Fig. 3(a)).[26] When computing an output, transformers ingest the entire input (*e.g.*, a sentence or an image) and use self-attention to decide which parts of the input are most relevant to each other. In other words, the model learns to weigh the influence of different input elements on each other dynamically. This attention-driven approach allows transformers to capture long-range dependencies in data effectively – for example, a word at the beginning of a sentence can directly influence the interpretation of a word at the end, because the model can access and evaluate both simultaneously. The ability to look at all parts of the input at once is also essential to mitigate issues including the vanishing-gradient problem that RNNs face when dealing with long sequences. Transformers have since become the dominant model in natural language processing and are increasingly used in other domains (with variants such as the Vision Transformers for images).[118]

The application of transformers has recently expanded to the design of metamaterials. In 2023, Chen *et al.* introduced the first encoder-only transformers for both forward and inverse design of broadband solar metamaterial absorbers (Fig. 3(b)).[28] The studied absorbers comprise 6 subwavelength layers with thicknesses ranging from 0 to 100 nm for Ge, Si, TiO_2, Al_2O_3, SiO_2 and 0–200 nm for MgF_2. To handle the high-dimensional spectral data, the input spectrum is segmented into multiple patches. Each patch is embedded using one-dimensional convolution before being fed into a transformer encoder. This segmentation and positional embedding help overcome overfitting and dimension mismatch issues,

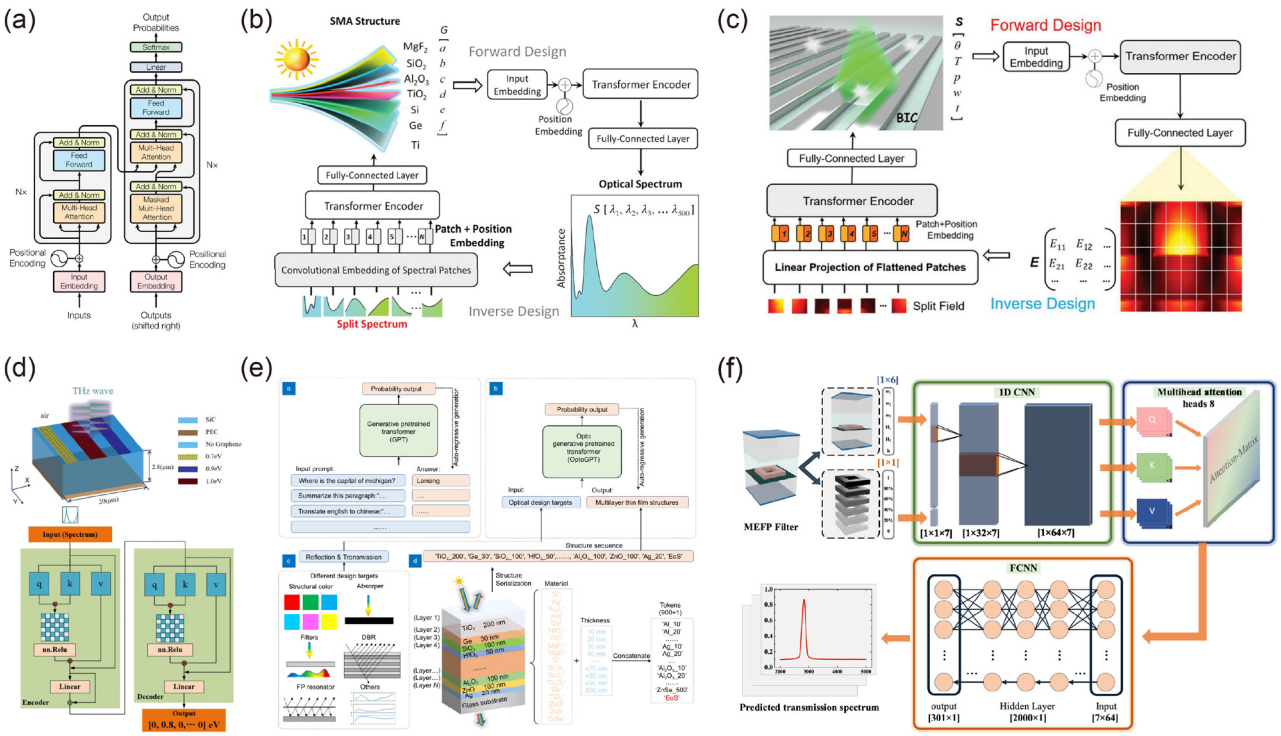

Fig. 3 Transformer and self-attention for AI-assisted design of metasurfaces. (a) Transformer architecture. Reproduced from ref. 26 with permission from Google, which grants reproduction of tables and figures for scholarly works provided proper attribution is given. (b) Encoder-only transformers for the design of broadband solar metamaterial absorbers. Reproduced from ref. 28 Copyright 2023. The authors, Advanced Photonics Research published by Wiley-VCH GmbH, under the terms of the Creative Commons Attribution License https://creativecommons.org/licenses/by/4.0/. (c) Dielectric metasurface design based on encoder-only transformer models. Reproduced from ref. 119 Copyright 2023. The authors, Advanced Optical Materials published by Wiley-VCH GmbH. (d) Improved transformer combined with a CGAN for the inverse-design of graphene terahertz multi-resonant metasurfaces. Reproduced with permission from ref. 120. Copyright 2023 IEEE. (e) GPT for inverse design of multilayer thin film structures. Reprinted with permission from ref. 121. The article is licensed under a CC-BY 4.0 License https://creativecommons.org/licenses/by/4.0/. (f) Mid-infrared metasurface-embedded Fabry–Perot filters design *via* FC layers that couple with a CNN-self-attention module. Reproduced with permission from ref. 122. Copyright 2024 American Chemical Society.

enabling the network to learn the underlying physical relationships effectively, and is expected to be implemented frequently in future works. After training, prediction is millisecond-scale on one NVIDIA GeForce GTX 3080 Ti. This work represents a significant breakthrough in AI-assisted design of metamaterials incorporating transformers, a new paradigm in navigating complex optical design spaces with unprecedented efficiency and accuracy. In the next study, Chen *et al.* (2023) have extended their approach to design a dielectric metasurface where the incident angle is tuned to mediate the coupling between guided-mode resonances and quasi-BICs (Fig. 3(c)).[119] With unit cells characterized by 5 parameters and a training set containing 23 681 simulation data points, their encoder-only transformers achieved an MSE of 4.56×10^{-3}, which represents a 17.6% improvement over a FC layers network, and a 34.4% reduction in training parameters. Other encoder-only approaches have also emerged. For example, Niu *et al.* (2023) have employed a VAE with an encoder-only transformer for the inverse design of metasurfaces,[123] while Yin *et al.* (2025) have combined a transformer encoder for forward prediction (~17 ms per prediction) with a visual attention network for inverse design.[124] In another study, Ma *et al.* (2025) have integrated FC layers with a transformer to achieve the inverse design of a metasurface absorber.[125] These models leverage the transformer encoder's ability to capture long-range dependencies within metamaterial data in parallel, making them well-suited for both forward and inverse tasks.

Some studies have implemented other transformer architectures.[126] For instance, Huang *et al.* (2024) have applied an improved transformer combined with a CGAN for the inverse-design of graphene terahertz multi-resonant metasurfaces, represented by a 20-element chemical potential vector (Fig. 3(d)).[120] After trained on 19 000 data points, the network achieved a test accuracy of 96.14% compared to 94.27% for an FC-layer model. In a separate effort, Ma *et al.* (2024) implemented a decoder-only transformer, which is also known as a Generative Pre-trained Transformer (GPT), for inverse design of multilayer thin film structures (Fig. 3(e)).[121] Rather than fixing the material for each layer and optimizing only the thickness, they introduced a "structure token" that defines both the material and its thickness, such as "Al_10". This approach not only overcomes the limitations of fixed output sizes but also enables handling diverse design scenarios, including variable numbers of layers, distinct material combinations, and different incidence angles and polarization states.

Despite these advances, transformers generally require large data sets because of their limited inductive biases.[127] Early studies noted that RNN-based sequence-to-sequence models could match or exceed the performance of transformers on small parallel data sets, underscoring the training challenges when data is limited.[128] More recent work has demonstrated that careful hyperparameter tuning, regularization, and strategies such as reducing network depth, employing smaller token vocabularies, or leveraging pre-training and transfer learning can help transformers perform in low-resource settings.[128] Therefore, we believe that relaxing the data requirements of transformers remains an important area for future research. Notwithstanding the remarkable achievements of Vision Transformers in computer vision tasks, their potential for image-based metamaterial design remains largely untapped, presenting a promising avenue for future research.

Beyond transformer architectures, incorporating self-attention into alternative network structures has also boosted the efficiency of metamaterial design. For example, Zeng *et al.* (2023) have investigated the data shift in electromagnetic solvers for 1D grating couplers by integrating a ResNet with mixed training and multihead attention.[129] They observed that when models are trained on data based on randomly generated nano-structures but then applied to predict optimized designs, a mismatch in data distributions (data shift) causes a significant drop in prediction accuracy. They also introduced a mixed training strategy, where a small fraction of the optimized (shifted) data is blended into the training set. The reported model achieved an MSE of 1.35×10^{-4} for coupling efficiency prediction, marking an improvement by 167% compared to a ResNet without attention. Similarly, Yuan *et al.* (2024) have inversely-designed tunable mid-infrared metasurface-embedded Fabry–Perot filters *via* FC layers that couple with a CNN-self-attention module for forward modelling (Fig. 3(f)).[122] Trained on 19 473 simulation samples, this approach yielded a test R^2 of 0.973 for predicting transmission as defined in their paper, approximately 10% higher than a comparable network without self-attention. These studies demonstrate the potential of self-attention mechanisms in the design of metamaterials. In particular, the reported results confirm that integrating self-attention mechanisms can reduce prediction errors and enhance model accuracy. Further investigation into such integrations, which lead to systems typically requiring less training data than transformers, is expected to advance the overall capabilities of AI-assisted metamaterial design.

Prediction of mutual coupling effects

Mutual coupling refers to the complex electromagnetic interactions between closely spaced meta-atoms, which can significantly alter the idealized responses assumed during conventional design.[130] Traditional design methods often apply periodic boundary conditions or approximations that neglect near-field interactions, which can lead to discrepancies between unit-cell based simulation results and actual device performance, particularly for phase-sensitive metasurfaces utilized for wavefront manipulation.[131] Modelling these interactions using full-wave electromagnetic simulations demands extensive computational resources, thereby hindering efficient design optimization. AI-based methods address this challenge by learning the mapping between meta-atom geometry, local environment, and optical response.[114,132–135]

In 2021, An and colleagues developed a CNN to accurately predict the electromagnetic responses of individual meta-atoms when mutual coupling between nonidentical neighbours is present (Fig. 4(a)).[136] The core idea involved translating the physical configurations of the target and adjacent meta-atoms into high-resolution, binarized images, where

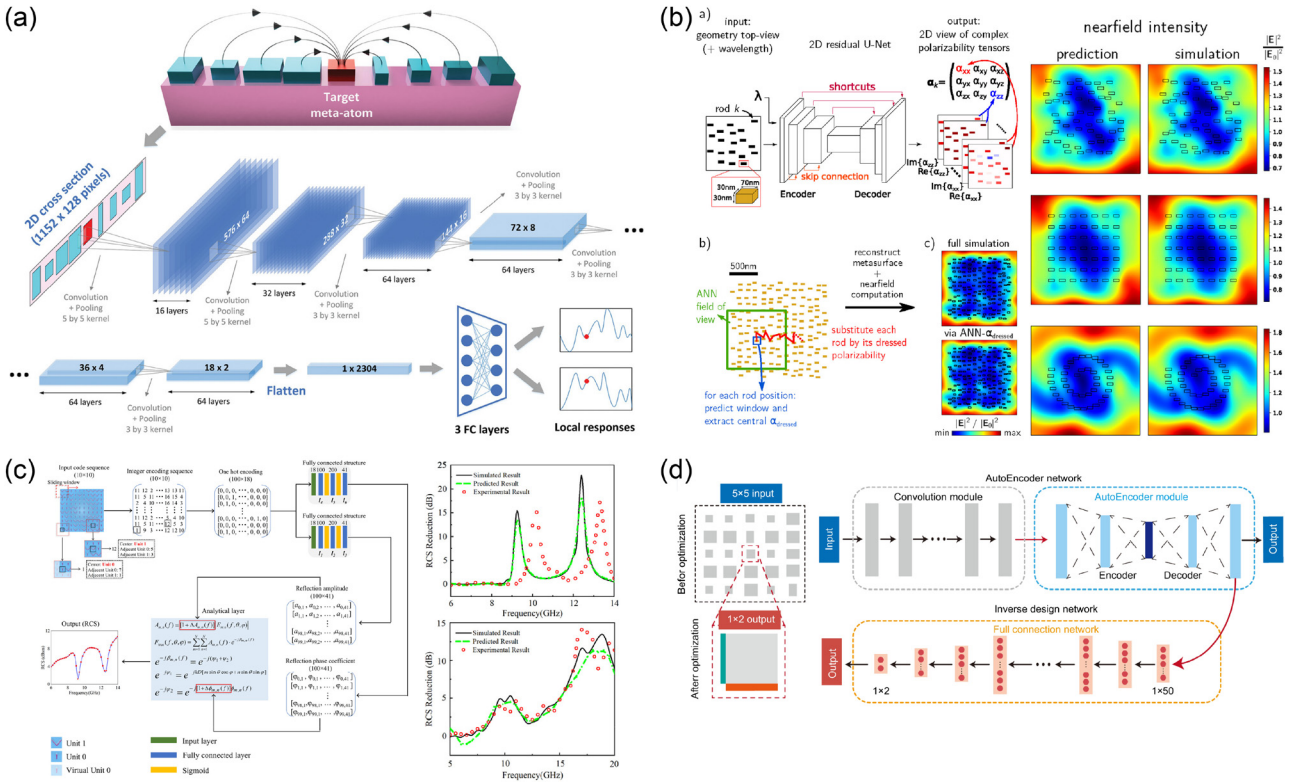

Fig. 4 AI-assisted design of metasurfaces with mutual coupling effects. (a) Prediction of electromagnetic responses of individual meta-atoms when mutual coupling between nonidentical neighbours is present *via* CNN. Reproduced from ref. 136 Copyright 2021. The authors, Advanced Optical Materials published by Wiley-VCH GmbH (b) U-Net-based CNN for the modelling of complex, aperiodic plasmonic metasurfaces that can extend to arbitrarily large sizes. Reproduced with permission from ref. 53. Copyright 2022 American Chemical Society. (c) Rapid calculation and optimization of metasurfaces incorporating meta-atom interactions. Reproduced from ref. 137. Copyright 2023. The authors, Advanced Photonics Research published by Wiley-VCH GmbH, under the terms of the Creative Commons Attribution License https://creativecommons.org/licenses/by/4.0/. (d) A DL optimizer for large-aperture meta-lens design *via* AE. Reprinted with permission from ref. 138. The article is licensed under a CC-BY 4.0 License. https://creativecommons.org/licenses/by/4.0.

dielectric regions and voids were distinctly marked, then processing these images through the CNN to extract key spatial features. When integrated with a global optimization, the method increased a beam deflection efficiency from 41.3% to 68.8% and improved a meta-lens's focusing efficiency by over 20%. In addition, An *et al.* demonstrated that by accounting for coupling perturbations, devices such as beam deflectors and meta-lenses exhibit significantly enhanced efficiency, ensuring that a larger fraction of the incident energy is effectively manipulated. In 2022, Majorel *et al.* developed a U-Net-based CNN to model optical responses of complex, aperiodic plasmonic metasurfaces that can extend to arbitrarily large sizes (Fig. 4(b)).[53] Instead of repeatedly calculating detailed optical interactions for every configuration, the authors proposed to approximate a "dressed polarizability" for each nanostructure. This quantity encapsulates how local interactions (due to neighboring coupling, substrates, and other environmental factors) alter the response of an individual nanostructure. The CNN takes a 2D top-view image of the nanostructure arrangement along with wavelength information as input and is trained to output the complex-valued dressed polarizability tensor for each nanostructure. This method provided scalability to arbitrarily large and complex geometries for future references. In 2023, Ma *et al.* proposed a DL model for rapid calculation and optimization of metasurfaces incorporating meta-atom interactions by training separate network blocks to associate reflection phase and amplitude with specific meta-atoms (Fig. 4(c)).[137] Rather than treating the DNN as a black box, their approach interprets weight values in cascaded dense layers as representing physical mechanisms of electromagnetic scattering. In the same year, Ha *et al.* developed a DL optimizer for large-aperture meta-lens design that segments the lens into overlapping 5 × 5 super meta-atoms to capture local lattice interactions (Fig. 4(d)).[138] Their architecture integrates an AE to extract low-dimensional representations of geometrical features with an inverse-design network that refines meta-atom dimensions to mitigate coupling-induced phase errors, resulting in a fabricated meta-lens with a 1 mm radius and a relative focusing efficiency of 93.4% (compared to the ideal focusing efficiency).

Moving forward, research on AI-assisted metasurface design is expected to further embrace models that inherently capture meta-atom responses over varying length scales while maintaining physical interpretability. Emerging network architec-

tures such as transformers offer the ability to learn long-range dependencies across an entire metasurface, using self-attention or fixed-attention to account for both local and global coupling effects without a preset neighbor limit. Equally important is infusing more physics knowledge into the learning process to avoid purely black-box behaviour. This could involve embedding constraints like energy conservation, reciprocity, or known coupling formulas into model architectures or loss functions, as well as designing network outputs that correspond to physically meaningful parameters.

Robust and fabrication-friendly metamaterial design

DNNs are making otherwise intractable problems a reality. In inverse-design, one typically seeks to maximize a set of performance criteria given a range of input parameter values. While these ranges may correspond to available material values or geometrical dimensions, such constraints are not the same as tolerances. In fact, tolerances are usually not considered *in situ* during optimization, but rather *a posteriori*, if at all. This means that the optimizer has no idea about the sensitivity of the response surface (*i.e.*, the hyper-dimensional objective space) to changes in input parameters. It is actually quite possible that the optimizer, simply seeking to maximize nominal performance, finds a solution that is highly-performant, but quite sensitive to input uncertainties. Designers therefore need to have previous experience when analysing optimized designs to anticipate their potential sensitivity. Otherwise, tolerance analysis may be performed using conventional Monte Carlo methods to estimate a design's guaranteed minimum performance given a set of input tolerances/uncertainties. However, in the case of computationally expensive models such as those requiring full-wave analysis or those with many input dimensions, this kind of analysis itself can be prohibitive. To this end, engineers have used surrogate modelling approaches based on radial basis functions or the Kriging model to help with such analysis. To this end, Easum *et al.* (2018)[139] introduced an optimization algorithm that iteratively trains surrogate models which accurately capture response surface features in order to calculate a design's "tolerance hypervolume". Due to the multi-objective nature of the algorithm, it presents designers with a Pareto front of optimized designs that showcase the trade-offs between performance objectives and robustness. While this technique represented a breakthrough in RF antenna optimization, classical surrogate models aren't equipped to handle the more sophisticated nonlinear relationships seen in nanophotonic meta-devices. Therefore, new techniques were needed in order to capture design robustness in freeform optical meta-devices.

Wen *et al.* demonstrated in 2020 how a progressively growing GAN (PGGAN) can learn how to output highly efficient and robust metasurfaces using only a sparse training data set (Fig. 5(a)).[140] The progressive growth aspect of the network enabled more robust learning of local topological features in the training set while a self-attention mechanism allowed the network to capture global features. The PGGAN was able to significantly accelerate the process for producing robust optimizes designs compared to the conventional topology (*i.e.*, adjoint) optimization process. In 2021, Jenkins *et al.* demonstrated a U-Net based architecture that achieved extremely accurate prediction of a metasurface's performance under geometric variations (Fig. 5(b)).[141] The network was used in conjunction with a more conventional multi-objective optimization algorithm in order to quantify both nominal (*i.e.*, no variations) and guaranteed minimum performances over all possible geometric deviations. Computing the guaranteed minimum performance is difficult as it requires an exhaustive evaluation of all possible variations; the minimum performance does not always occur at the extremes of the uncertainty range. Using this approach, the authors presented a result that trades off a few percent of nominal performance for over a 100% increase in guaranteed minimum performance. Moreover, the hybrid DL approach reduced the inverse-design process from a potential many months' time scale to that of just a few days, overall speedups 14.8 times for a single optimization ignoring startup time and 4.37 times including it.

In 2023, Tanriover *et al.* combined an AE and FC NN to produce a DL model capable of generating manufacturable freeform dielectric meta-atoms (Fig. 5(c)).[70] Their approach sought to improve upon model generalizability and fabrication feasibility compared to other solutions. The forward model exhibited generalizability in material dispersion, source polarization, and wavelength range of operation and was subsequently connected to a GA to perform design optimization. In the same year, Ueno *et al.* presented a DL framework for producing fabrication-friendly metasurfaces. The authors demonstrated a dual-band optical collimator whose inverse-design was accelerated by the DNN (Fig. 5(d)).[142] The free-form meta-atoms were generated using a predictive neural network (PNN) which was trained to accurately predict the transmission phase and amplitude of candidate designs.

Recent advancements and perspectives

Recent research highlights several emerging trends that are shaping the future of AI-assisted metamaterial design. These approaches aim to overcome current limitations (such as data scarcity, limited generalization, or design complexity) and open new possibilities for metamaterial engineering.

One emerging approach is the hybridization of different neural network types within a single design framework. The motivation is that complex metamaterial design tasks often involve multiple representations (*e.g.* geometric patterns, spectral responses, parametric features) that may be best handled by different network architectures working in concert. For example, Chen *et al.* (2025) reported a CNN-LSTM-A model combining convolutional layers, LSTM networks, and attention mechanisms that achieved a prediction accuracy of 0.993 for

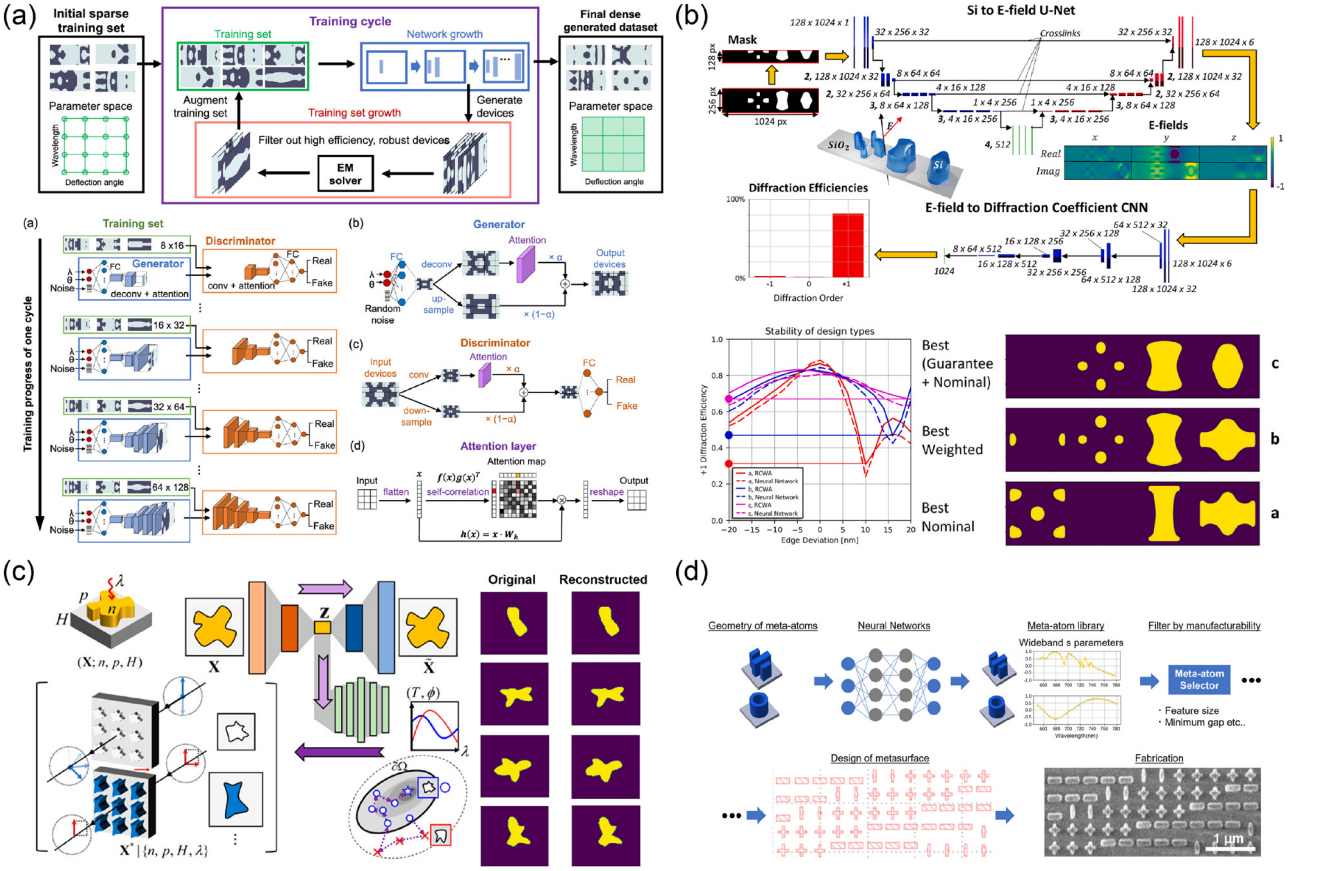

Fig. 5 AI-assisted design of robust and fabrication-friendly metasurfaces. (a) PGGAN with self-attention rapidly output freeform metasurface designs that surpass topology-optimized devices in efficiency and robustness. Reproduced with permission from ref. 140. Copyright 2020 American Chemical Society. (b) A U-net based DNN with evolutionary optimization to design metasurfaces whose efficiency persists across fine-grained fabrication-induced edge deviations. Reprinted with permission from ref. 141. The article is licensed under a Creative Commons Attribution 4.0 International License https://creativecommons.org/licenses/by/4.0/. (c) An end-to-end generative-modelling pipeline that learns manufacturable free-form dielectric metasurface shapes. Reproduced with permission from ref. 70. Copyright 2022 American Chemical Society. (d) A DL-generated, fabrication-constrained library of free-form *meta*-atoms for the design of metasurface collimators. Reprinted with permission from ref. 142. The article is licensed under a Creative Commons Attribution 4.0 International License https://creativecommons.org/licenses/by/4.0/.

the spectral response of all-dielectric trimer metasurfaces exhibiting double Fano resonances.[143] We believe that these hybrid network approaches are improving the stability and fidelity of AI-driven design, which can further offer new possibilities for designing complex metamaterials. Additionally, integrating large language models appears promising because, as Kim *et al.* (2025), Zhang *et al.* (2025) and Lu *et al.* (2025) demonstrated, it lets researchers achieve comparable results with less ML expertise and less code.[144–146]

Another major trend is the integration of physical laws and domain knowledge directly into DL models, a practice known as PINNs. By embedding principles like Maxwell's equations or partial differential equations (PDEs) into the training process or network architecture, these PINNs can significantly reduce the need for large training datasets and improve model reliability. For instance, a convolution-based PINN with U-Net backbones accurately simulates near-field and far-field responses with speed improvements of up to 10 000× over traditional solvers.[147] Ongoing works may seek to extend this framework by incorporating multiple physical domains (electromagnetic, thermal, mechanical) into AI models. This trend is expected to grow as it reduces the need for massive training datasets, but the long training time and high GPU memory consumption represent major obstacles for this method.

Newer and complex AI architectures are showing an ever-growing potentiality. A refined GAN paired with an agent model based on the Swin Transformer[148] has enabled efficient generation of metasurface patterns from spectral data, achieving MSE as low as 6×10^{-3} between simulated and generated spectra.[149] Likewise, a U-net with CGAN framework has facilitated forward and inverse design of terahertz metasurfaces with multifunctional responses, achieving inverse prediction accuracies exceeding 94%.[150] While these sophisticated architectures yield performance improvements, it's important to balance these gains against the increased computational costs, extended training times they entail, and a potential greater demand for training data sets. In some cases, a simpler model may offer a more efficient solution for less complex design challenges.

In addition to the supervised methods, unsupervised learning offers an alternative framework for metamaterial design. One approach uses the K-Nearest Neighbor (KNN) algorithm, which requires fewer data and computational resources than NNs. KNN clusters and interpolates metamaterial configurations to yield new geometries with defined property combinations without direct supervision. Recently, Fan *et al.* (2025) reported an inverse design method for optical power splitters that combined KNN with particle swarm optimization.[151] This unsupervised learning method offers a fresh perspective on the inverse design of photonics. We believe that unsupervised learning models can be further extended to other metamaterial design challenges.

Conclusions

In this review, we have provided a comprehensive overview of data-driven approaches in nanophotonics, with a particular focus on the design and optimization of AI-enabled metadevices. Our discussion highlighted the significant strides achieved through both image-based and parameter-based DL methods. Advanced techniques including CNNs, RNNs, GANs, VAEs, and most recently, transformers have demonstrated their ability to efficiently navigate complex, high-dimensional design spaces and account for intricate physical phenomena such as mutual coupling effects. These methodologies not only overcome the computational limitations of traditional design approaches but also enable the rapid prediction and inverse design of multifunctional photonic architectures.

To conclude, the fusion of artificial intelligence with nanophotonic engineering marks a transformative shift in metadevices development. As described throughout this review, AI-driven strategies hold immense promise for enhancing design precision, accelerating optimization processes, and ultimately facilitating the development of next-generation photonic platforms. Looking ahead, future research is expected to delve deeper into hybrid network architectures, physics-informed learning, and attention-based models, thereby broadening the scope and impact of data-driven nanophotonic design.

Conflicts of interest

There are no conflicts to declare.

Data availability

No data were used for the research described in the article.

Acknowledgements

This work was supported by the John L. and Genevieve H. McCain endowed chair professorship at The Pennsylvania State University.

References

1 N. I. Zheludev and Y. S. Kivshar, *Nat. Mater.*, 2012, **11**, 917–924.
2 W. Cai, U. K. Chettiar, H.-K. Yuan, V. C. De Silva, A. V. Kildishev, V. P. Drachev and V. M. Shalaev, *Opt. Express*, 2007, **15**, 3333.
3 J. Valentine, S. Zhang, T. Zentgraf, E. Ulin-Avila, D. A. Genov, G. Bartal and X. Zhang, *Nature*, 2008, **455**, 376–379.
4 J. K. Gansel, M. Wegener, S. Burger and S. Linden, *Opt. Express*, 2010, **18**, 1059.
5 H.-T. Chen, A. J. Taylor and N. Yu, *Rep. Prog. Phys.*, 2016, **79**, 076401.
6 N. Yu, P. Genevet, M. A. Kats, F. Aieta, J.-P. Tetienne, F. Capasso and Z. Gaburro, *Science*, 2011, **334**, 333–337.
7 L. Huang, X. Chen, H. Mühlenbernd, G. Li, B. Bai, Q. Tan, G. Jin, T. Zentgraf and S. Zhang, *Nano Lett.*, 2012, **12**, 5750–5755.
8 L. Huang, S. Zhang and T. Zentgraf, *Nanophotonics*, 2018, **7**, 1169–1190.
9 H. Zhou, C. Zhao, C. He, L. Huang, T. Man and Y. Wan, *Nanophotonics*, 2024, **13**, 419–441.
10 Y. Hao and R. Mittra, *FDTD Modeling of Metamaterials: Theory and Applications*, Artech House, Boston London, 2009.
11 J.-M. Jin, *The Finite Element Method in Electromagnetics*, IEEE Press, Piscataway, NJ, 3rd edn, 2014.
12 Y. Dong, S. An, H. Jiang, B. Zheng, H. Tang, Y. Huang, H. Zhao and H. Zhang, *Prog. Quantum Electron.*, 2025, 100554.
13 S. D. Campbell, D. Sell, R. P. Jenkins, E. B. Whiting, J. A. Fan and D. H. Werner, *Opt. Mater. Express*, 2019, **9**, 1842.
14 M. M. R. Elsawy, S. Lanteri, R. Duvigneau, J. A. Fan and P. Genevet, *Laser Photonics Rev.*, 2020, **14**, 1900445.
15 S. J. Russell and P. Norvig, Artificial intelligence: a modern approach, Pearson, Hoboken, 4th edn, 2021.
16 J. Song, J. Lee, N. Kim and K. Min, *Int. J. Precis. Eng. Manuf.*, 2024, **25**, 225–244.
17 I. Malkiel, A. Nagler, M. Mrejen, U. Arieli, L. Wolf and H. Suchowski, *arXiv*, 2017, preprint, arXiv.1702.07949, DOI: **10.48550/arXiv.1702.07949**.
18 M. H. Tahersima, K. Kojima, T. Koike-Akino, D. Jha, B. Wang, C. Lin and K. Parsons, *Sci. Rep.*, 2019, **9**, 1368.
19 W. Ma, F. Cheng and Y. Liu, *ACS Nano*, 2018, **12**, 6326–6334.
20 J. Peurifoy, Y. Shen, L. Jing, Y. Yang, F. Cano-Renteria, B. G. DeLacy, J. D. Joannopoulos, M. Tegmark and M. Soljačić, *Sci. Adv.*, 2018, **4**, eaar4206.
21 S. Hochreiter and J. Schmidhuber, *Neural Comput.*, 1997, **9**, 1735–1780.
22 W. Deng, Z. Xu, J. Wang and J. Lv, *Opt. Lett.*, 2022, **47**, 3239.
23 P. Pillai, P. Pal, R. Chacko, D. Jain and B. Rai, *Sci. Rep.*, 2021, **11**, 18629.

24 Q. Li, H. Fan, Y. Bai, Y. Li, M. Ikram, Y. Wang, Y. Huo and Z. Zhang, *New J. Phys.*, 2022, **24**, 063005.
25 S. An, B. Zheng, M. Y. Shalaginov, H. Tang, H. Li, L. Zhou, J. Ding, A. M. Agarwal, C. Rivero-Baleine, M. Kang, K. A. Richardson, T. Gu, J. Hu, C. Fowler and H. Zhang, *Opt. Express*, 2020, **28**, 31932.
26 A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, *arXiv*, 2017, preprint, arXiv.1706.03762, DOI: **10.48550/arXiv.1706.03762**.
27 OpenAI, *et al.*, *arXiv*, 2023, preprint, arXiv.2303.08774, DOI: **10.48550/arXiv.2303.08774**.
28 W. Chen, Y. Gao, Y. Li, Y. Yan, J. Ou, W. Ma and J. Zhu, *Adv. Sci.*, 2023, **10**, 2206718.
29 A. Ueno, J. Hu and S. An, *npj Nanophotonics*, 2024, **1**, 36.
30 A. Khaireh-Walieh, D. Langevin, P. Bennet, O. Teytaud, A. Moreau and P. R. Wiecha, *Nanophotonics*, 2023, **12**, 4387–4414.
31 T. Asano and S. Noda, *Opt. Express*, 2018, **26**, 32704.
32 I. Sajedian, J. Kim and J. Rho, *Microsyst. Nanoeng.*, 2019, **5**, 27.
33 J.-F. Masson, J. S. Biggins and E. Ringe, *Nat. Nanotechnol.*, 2023, **18**, 111–123.
34 M. K. Chen, X. Liu, Y. Sun and D. P. Tsai, *Chem. Rev.*, 2022, **122**, 15356–15413.
35 Y. Fu, X. Zhou, Y. Yu, J. Chen, S. Wang, S. Zhu and Z. Wang, *Nanophotonics*, 2024, **13**, 1239–1278.
36 Y. Xu, X. Zhang, Y. Fu and Y. Liu, *Photonics Res.*, 2021, **9**, B135.
37 Y. Jin, L. He, Z. Wen, B. Mortazavi, H. Guo, D. Torrent, B. Djafari-Rouhani, T. Rabczuk, X. Zhuang and Y. Li, *Nanophotonics*, 2022, **11**, 439–460.
38 K. Yao, R. Unni and Y. Zheng, *Nanophotonics*, 2019, **8**, 339–366.
39 R. S. Hegde, *Nanoscale Adv.*, 2020, **2**, 1007–1023.
40 P. R. Wiecha, A. Arbouet, C. Girard and O. L. Muskens, *Photonics Res.*, 2021, **9**, B182.
41 W. Ma, Z. Liu, Z. A. Kudyshev, A. Boltasseva, W. Cai and Y. Liu, *Nat. Photonics*, 2021, **15**, 77–90.
42 D. Piccinotti, K. F. MacDonald, S. A. Gregory, I. Youngs and N. I. Zheludev, *Rep. Prog. Phys.*, 2021, **84**, 012401.
43 E. Tezsezen, D. Yigci, A. Ahmadpour and S. Tasoglu, *ACS Appl. Mater. Interfaces*, 2024, **16**, 29547–29569.
44 Q. Wang, M. Makarenko, A. Burguete Lopez, F. Getman and A. Fratalocchi, *Nanophotonics*, 2022, **11**, 2483–2505.
45 J. Jiang, M. Chen and J. A. Fan, *Nat. Rev. Mater.*, 2020, **6**, 679–700.
46 C. Qian, L. Tian and H. Chen, *Light: Sci. Appl.*, 2025, **14**, 93.
47 S. D. Campbell, R. P. Jenkins, P. J. O'Connor and D. Werner, *IEEE Antennas Propag. Mag.*, 2021, **63**, 16–27.
48 J. Park, S. Kim, D. W. Nam, H. Chung, C. Y. Park and M. S. Jang, *Nanophotonics*, 2022, **11**, 1809–1845.
49 D. Midtvedt, V. Mylnikov, A. Stilgoe, M. Käll, H. Rubinsztein-Dunlop and G. Volpe, *Nanophotonics*, 2022, **11**, 3189–3214.
50 W. Ji, J. Chang, H.-X. Xu, J. R. Gao, S. Gröblacher, H. P. Urbach and A. J. L. Adam, *Light: Sci. Appl.*, 2023, **12**, 169.
51 D. Lee, W. (Wayne) Chen, L. Wang, Y. Chan and W. Chen, *Adv. Mater.*, 2024, **36**, 2305254.
52 W. Liu, X. Wang and M. Zeng, *Opt. Lett.*, 2022, **47**, 5112.
53 C. Majorel, C. Girard, A. Arbouet, O. L. Muskens and P. R. Wiecha, *ACS Photonics*, 2022, **9**, 575–585.
54 S. Singh, R. Kumar, S. S. Panda and R. S. Hegde, *Digital Discovery*, 2024, **3**, 1612–1623.
55 Y. Teng, C. Li, S. Li, Y. Xiao and L. Jiang, *Opt. Laser Technol.*, 2023, **160**, 109058.
56 X. Han, Z. Fan, Z. Liu, C. Li and L. J. Guo, *InfoMat*, 2021, **3**, 432–442.
57 Y. Li, Y. Zhang, Y. Wang, J. Li, X. Jiang, G. Yang, K. Zhang, Y. Yuan, J. Fu, X. Di and C. Wang, *Adv. Opt. Mater.*, 2024, **12**, 2302657.
58 R. Zhu, T. Qiu, J. Wang, S. Sui, Y. Li, M. Feng, H. Ma and S. Qu, *J. Phys. D:Appl. Phys.*, 2020, **53**, 455002.
59 R. Zhu, T. Qiu, J. Wang, S. Sui, C. Hao, T. Liu, Y. Li, M. Feng, A. Zhang, C.-W. Qiu and S. Qu, *Nat. Commun.*, 2021, **12**, 2974.
60 W. Ma, F. Cheng, Y. Xu, Q. Wen and Y. Liu, *Adv. Mater.*, 2019, **31**, 1901111.
61 L. Zhu, W. Hua, C. Lv and Y. Liu, *J. Lightwave Technol.*, 2024, **42**, 5269–5278.
62 Y.-F. Liu, L.-Y. Xiao, W. Shao, L. Peng and Q. H. Liu, *IEEE Antennas Wirel. Propag. Lett.*, 2024, **23**, 4568–4572.
63 T. Qu, L. Zhu and Z. An, *Opt. Lett.*, 2023, **48**, 448.
64 Y. Li, Y. Wang, S. Qi, Q. Ren, L. Kang, S. D. Campbell, P. L. Werner and D. H. Werner, *IEEE Access*, 2020, **8**, 139983–139993.
65 S. An, B. Zheng, H. Tang, M. Y. Shalaginov, L. Zhou, H. Li, M. Kang, K. A. Richardson, T. Gu, J. Hu, C. Fowler and H. Zhang, *Adv. Opt. Mater.*, 2021, **9**, 2001433.
66 R. Yu, Y. Liu and L. Zhu, *Opt. Express*, 2022, **30**, 35776.
67 Z. Zhang, C. Yang, Y. Qin, H. Feng, J. Feng and H. Li, *Nanophotonics*, 2023, **12**, 3871–3881.
68 H. Yang, C. Yang and H. Li, *ACS Photonics*, 2025, **12**, 1184–1195.
69 T. Gahlmann and P. Tassin, *Phys. Rev. B*, 2022, **106**, 085408.
70 I. Tanriover, D. Lee, W. Chen and K. Aydin, *ACS Photonics*, 2022, 875–883.
71 K. Cho, B. van Merrienboer, D. Bahdanau and Y. Bengio, *arXiv*, 2014, preprint, arXiv:1409.1259, DOI: **10.48550/arXiv:1409.1259**.
72 I. Malkiel, M. Mrejen, A. Nagler, U. Arieli, L. Wolf and H. Suchowski, *Light: Sci. Appl.*, 2018, **7**, 60.
73 C. C. Nadell, B. Huang, J. M. Malof and W. J. Padilla, *Opt. Express*, 2019, **27**, 27523.
74 D. Xu, Y. Luo, J. Luo, M. Pu, Y. Zhang, Y. Ha and X. Luo, *Opt. Mater. Express*, 2021, **11**, 1852.
75 X. Liao, L. Gui, Z. Yu, T. Zhang and K. Xu, *Opt. Mater. Express*, 2022, **12**, 758.
76 H. Zhang, L. Kang, S. D. Campbell, K. Zhang, D. H. Werner and Z. Cao, *Opt. Express*, 2025, **33**, 18928.

77 D. Liu, Y. Tan, E. Khoram and Z. Yu, *ACS Photonics*, 2018, **5**, 1365–1369.
78 A. Mall, A. Patil, D. Tamboli, A. Sethi and A. Kumar, *J. Phys. D:Appl. Phys.*, 2020, **53**, 49LT01.
79 J. Hou, H. Lin, W. Xu, Y. Tian, Y. Wang, X. Shi, F. Deng and L. Chen, *IEEE Access*, 2020, **8**, 211849–211859.
80 J. H. Han, Y.-C. Lim, R. M. Kim, J. Lv, N. H. Cho, H. Kim, S. D. Namgung, S. W. Im and K. T. Nam, *ACS Nano*, 2023, **17**, 2306–2317.
81 C. Luo, T. Sang, Z. Ge, J. Lu and Y. Wang, *Opt. Express*, 2024, **32**, 13978.
82 S. An, C. Fowler, B. Zheng, M. Y. Shalaginov, H. Tang, H. Li, L. Zhou, J. Ding, A. M. Agarwal, C. Rivero-Baleine, K. A. Richardson, T. Gu, J. Hu and H. Zhang, *ACS Photonics*, 2019, **6**, 3196–3207.
83 L. Gao, X. Li, D. Liu, L. Wang and Z. Yu, *Adv. Mater.*, 2019, **31**, 1905467.
84 K.-F. Lin, C.-C. Hsieh, S.-C. Hsin and W.-F. Hsieh, *Appl. Opt.*, 2019, **58**, 8914.
85 X. Li, J. Shu, W. Gu and L. Gao, *Opt. Mater. Express*, 2019, **9**, 3857.
86 I. Sajedian, H. Lee and J. Rho, *Sci. Rep.*, 2019, **9**, 10899.
87 R. Unni, K. Yao and Y. Zheng, *ACS Photonics*, 2020, **7**, 2703–2712.
88 I. Tanriover, W. Hadibrata and K. Aydin, *ACS Photonics*, 2020, **7**, 1957–1964.
89 C. Qiu, X. Wu, Z. Luo, H. Yang, G. Wang, N. Liu and B. Huang, *Opt. Commun.*, 2021, **483**, 126641.
90 L. Xu, M. Rahmani, Y. Ma, D. A. Smirnova, K. Z. Kamali, F. Deng, Y. K. Chiang, L. Huang, H. Zhang, S. Gould, D. N. Neshev and A. E. Miroshnichenko, *Adv. Photonics*, 2020, **2**, 1.
91 R. Unni, K. Yao, X. Han, M. Zhou and Y. Zheng, *Nanophotonics*, 2021, **10**, 4057–4065.
92 A. Lininger, M. Hinczewski and G. Strangi, *ACS Photonics*, 2021, **8**, 3641–3650.
93 M. Zandehshahvar, Y. Kiarashi, M. Chen, R. Barton and A. Adibi, *Opt. Lett.*, 2021, **46**, 2634.
94 W. Huang, Z. Wei, B. Tan, S. Yin and W. Zhang, *J. Phys. D: Appl. Phys.*, 2021, **54**, 135102.
95 Z. Sun, B. Xu, F. Jin, G. Zhou and L. Lin, *IEEE J. Sel. Top. Quantum Electron.*, 2022, **28**, 1–9.
96 I. Tanriover, W. Hadibrata, J. Scheuer and K. Aydin, *Opt. Express*, 2021, **29**, 27219.
97 Y. Deng, S. Ren, K. Fan, J. M. Malof and W. J. Padilla, *Opt. Express*, 2021, **29**, 7526.
98 X. Xu, C. Sun, Y. Li, J. Zhao, J. Han and W. Huang, *Opt. Commun.*, 2021, **481**, 126513.
99 S. Noureen, M. Q. Mehmood, M. Ali, B. Rehman, M. Zubair and Y. Massoud, *Nanoscale*, 2022, **14**, 16436–16449.
100 F. Gao, Z. Zhang, Y. Xu, L. Zhang, R. Yan and X. Chen, *J. Opt. Soc. Am. B*, 2022, **39**, 1511.
101 R. Shen, R. He, L. Chen and J. Guo, *Opt. Mater. Express*, 2022, **12**, 3600.
102 H. Lin, J. Hou, J. Jin, Y. Wang, R. Tang, X. Shi, Y. Tian and W. Xu, *Opt. Express*, 2022, **30**, 3076.
103 R. Li, J. Cheng, X. Dong and S. Chang, *J. Phys. D:Appl. Phys.*, 2022, **55**, 155106.
104 T. Knightley, A. Yakovlev and V. Pacheco-Peña, *Adv. Opt. Mater.*, 2023, **11**, 2202351.
105 Y. Chen, Z. Ding, J. Wang, J. Zhou and M. Zhang, *Opt. Lett.*, 2022, **47**, 5092.
106 Y. Qiu, S. Chen, Z. Hou, J. Wang, J. Shen and C. Li, *Micromachines*, 2023, **14**, 789.
107 P. Liu, Y. Zhao, N. Li, K. Feng, S. G. Kong and C. Tang, *Opt. Lasers Eng.*, 2024, **174**, 107933.
108 Z. Jiang, Z. Gan, C. Liang and W.-D. Li, *Nanophotonics*, 2024, **13**, 1181–1189.
109 S. Yu, T. Zhang, J. Dai and K. Xu, *Opt. Express*, 2023, **31**, 39852.
110 T. Jahan, T. Dash, S. E. Arman, R. Inum, S. Islam, L. Jamal, A. A. Yanik and A. Habib, *Nanoscale*, 2024, **16**, 16641–16651.
111 Y. Chen, Q. Wang, D. Cui, W. Li, M. Shi and G. Zhao, *Opt. Commun.*, 2024, **569**, 130793.
112 Z.-D. Wang, Y.-L. Meng, Y. Li, H. Gao, T. Zhang, G.-M. Pan, J. Kang and C.-L. Zhan, *Opt. Commun.*, 2024, **573**, 130995.
113 L. Zhu, W. Du, L. Dong and J. Wei, *Phys. Scr.*, 2024, **99**, 036002.
114 L. Fan, Y. Yu, C. Gao, X. Qu and C. Zhou, *Opt. Lett.*, 2024, **49**, 4318.
115 Y. Liu, Q. Geng, W. Zhan and Z. Geng, *Eng. Appl. Artif. Intell.*, 2025, **144**, 110172.
116 Y. Yu, S. You, Y. Zhang, L. Wang, H. Duan, H. He, Y. Wang, S. Luo, J. Xu, J. Huang and C. Zhou, *Appl. Phys. Lett.*, 2025, **126**, 071704.
117 J. Chen, Z. Zhang, Z. Huang and K. Cui, *Appl. Phys. Lett.*, 2025, **126**, 051703.
118 K. Han, Y. Wang, H. Chen, X. Chen, J. Guo, Z. Liu, Y. Tang, A. Xiao, C. Xu, Y. Xu, Z. Yang, Y. Zhang and D. Tao, *IEEE Trans. Antennas Propag.*, 2023, **45**, 87–110.
119 W. Chen, Y. Li, Y. Liu, Y. Gao, Y. Yan, Z. Dong and J. Zhu, *Adv. Opt. Mater.*, 2024, **12**, 2301697.
120 Y. Huang, N. Feng and Y. Cai, *J. Lightwave Technol.*, 2024, **42**, 1518–1525.
121 T. Ma, H. Wang and L. J. Guo, *Opto-Electron. Adv.*, 2024, **7**, 240062–240062.
122 X. Yuan, Z. Wei, Q. Ma, W. Ding and J. Guo, *ACS Appl. Mater. Interfaces*, 2024, **16**, 26500–26511.
123 C. Niu, M. Phaneuf, T. Qiu and P. Mojabi, *IEEE Open J Antennas Propag*, 2023, **4**, 641–653.
124 S. Yin, H. Zhong, W. Huang and W. Zhang, *Opt. Laser Technol.*, 2025, **181**, 111684.
125 J. Ma, Z. Ma, M. Li, Y. Li, B. Tan and S. Ding, *Phys. Scr.*, 2025, **100**, 016003.
126 Y. Gao, W. Chen, F. Li, M. Zhuang, Y. Yan, J. Wang, X. Wang, Z. Dong, W. Ma and J. Zhu, *Adv. Sci.*, 2024, **11**, 2405750.
127 Y. Liu, E. Sangineto, W. Bi, N. Sebe, B. Lepri and M. Nadai, in *Advances in Neural Information Processing Systems*, ed. M. Ranzato, A. Beygelzimer, Y. Dauphin,

P. S. Liang and J. W. Vaughan, Curran Associates, Inc., 2021, vol. 34, pp. 23818–23830.
128 S. Lankford, H. Afli and A. Way, *arXiv*, 2024, preprint, arXiv:2403.01985, DOI: **10.48550/arXiv.2403.01985**.
129 Z. Zeng, L. Wang, Y. Wu, Z. Hu, J. Evans, X. Zhu, G. Ye and S. He, *Nanomaterials*, 2023, **13**, 2778.
130 N. Liu and H. Giessen, *Angew. Chem., Int. Ed.*, 2010, **49**, 9838–9852.
131 A. E. Olk and D. A. Powell, *Phys. Rev. Appl.*, 2019, **11**, 064007.
132 M. V. Zhelyeznyakov, S. Brunton and A. Majumdar, *ACS Photonics*, 2021, **8**, 481–488.
133 Y. Ma, Q. Luo, C. Zhang and G. Yang, *IEEE Trans. Antennas Propag.*, 2024, **72**, 8443–8451.
134 Q. Bao, D. Zhang, X. Liu, T. Ma and J.-J. Xiao, *Opt. Laser Technol.*, 2025, **183**, 112273.
135 M. Li, Y. Zhang and Z. Ma, *IEEE J. Multiscale Multiphysics Comput. Tech.*, 2023, **8**, 40–48.
136 S. An, B. Zheng, M. Y. Shalaginov, H. Tang, H. Li, L. Zhou, Y. Dong, M. Haerinia, A. M. Agarwal, C. Rivero-Baleine, M. Kang, K. A. Richardson, T. Gu, J. Hu, C. Fowler and H. Zhang, *Adv. Opt. Mater.*, 2022, **10**, 2102113.
137 Y. Ma, J. F. Kolb, A. A. Ihalage, A. S. Andy and Y. Hao, *Adv. Photonics Res.*, 2023, **4**, 2200099.
138 Y. Ha, Y. Luo, M. Pu, F. Zhang, Q. He, J. Jin, M. Xu, Y. Guo, X. Li, X. Li, X. Ma and X. Luo, *Opto-Electron. Adv.*, 2023, **6**, 230133–230133.
139 J. A. Easum, J. Nagar, P. L. Werner and D. H. Werner, *IEEE Trans. Antennas Propag.*, 2018, **66**, 6706–6715.
140 F. Wen, J. Jiang and J. A. Fan, *ACS Photonics*, 2020, **7**, 2098–2104.
141 R. P. Jenkins, S. D. Campbell and D. H. Werner, *Nanophotonics*, 2021, **10**, 4497–4509.
142 A. Ueno, H.-I. Lin, F. Yang, S. An, L. Martin-Monier, M. Y. Shalaginov, T. Gu and J. Hu, *Nanophotonics*, 2023, **12**, 3491–3499.
143 Y. Chen, C. Mao, M. Li, W. Li, M. Shi and Q. Wang, *Opt. Commun.*, 2025, **574**, 131218.
144 M. Kim, H. Park and J. Shin, *Nanophotonics*, 2025, **14**, 1273–1282.
145 D. Lu, Y. Deng, J. M. Malof and W. J. Padilla, *arXiv*, 2025, preprint, arXiv:2404.15458, DOI: **10.48550/arXiv.2404.15458**.
146 H. Zhang, L. Kang, S. D. Campbell and D. H. Werner, *Nanophotonics*, 2025, DOI: **10.1515/nanoph-2025-0343**.
147 V. Medvedev, A. Erdmann and A. Rosskopf, *Opt. Express*, 2025, **33**, 1371.
148 Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin and B. Guo, in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, Montreal, QC, Canada, 2021, pp. 9992–10002.
149 J. Wang, B. Yao, Y. Niu, J. Ma, Y. Wang, Z. Qu, J. Duan and B. Zhang, *Adv. Compos. Hybrid Mater.*, 2025, **8**, 94.
150 H. Xia, S.-L. Chen, Y. Wang, Y. Zhao, H. Jia, R. Yang and Y. Jay Guo, *Opt. Laser Technol.*, 2025, **181**, 112041.
151 H. Fan, J. Pan, Y. Wang, Z. Yuan, M. Cheng, Q. Yang, D. Liu and L. Deng, *Nanophotonics*, 2025, **14**, 169–181.