J. Fluid Mech. (2026), *vol.* 1031, P1, doi:10.1017/jfm.2026.11256

JFM PERSPECTIVES



Routes towards an effective AI in CFD: an epistemological and technical perspective

Michaël Bauerheim

ISAE-SUPAERO, 10 Av. Marc Pélegrin, Toulouse, France

Corresponding author: Michaël Bauerheim, michael.bauerheim@isae-supaero.fr

(Received 28 January 2026; revised 28 January 2026; accepted 28 January 2026)

The integration of Artificial Intelligence (AI) into computational science (CS) and computational fluid dynamics (CFD) has raised profound epistemological debates concerning the nature of knowledge and its effectiveness in science. A central question in this discourse is whether AI can rival, or potentially surpass, the effectiveness of traditional mathematical methods in addressing the intricate challenges of CFD. In this work, I examine the concept of effectiveness within this context, highlighting the fundamental epistemological distinctions between AI-driven approaches and classical mathematical techniques. First, this analysis identifies four foundational pillars of effectiveness (PoEs) in scientific methods: (i) symmetries, which impose internal structure and coherence; (ii) scale separation, allowing specific treatments for the different scales and their interactions; (iii) sparsity, which simplifies complexity and enhances explicability; and (iv) semantic significance, which fosters abstraction, reasoning and interpretability. Yet, unlike mathematics where rigour ensures credibility by default, AI methods raise additional concerns of robustness and trust. Therefore, beyond the four PoEs, I also discuss credibility as a complementary pillar essential for the adoption of AI in the CFD community. The next critical step is to assess whether, and to what extent, AI can emulate or even outperform the roles and functions traditionally fulfilled by mathematical models. I therefore systematically review if, and how, these four pillar of effectiveness can be applied to AI-based algorithms. I show that those pillars are actually declined in a succession of technical advances that have shown promising results when using AI in CFD.

Key words: computational methods, machine learning

1. Introduction

Artificial Intelligence (AI) emerged approximately 70 years ago as a result of humanity's enduring ambition to model and compute reasoning and logic. The Dartmouth Summer

© The Author(s), 2026. Published by Cambridge University Press.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

Research Project on Artificial Intelligence, in 1956, marked a pivotal moment, operating under the conjecture that 'every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it' (McCarthy *et al.* 1956). This project catalysed a prolific wave of research aimed at replicating human capabilities, such as reasoning, vision (Russakovsky *et al.* 2015) and speech recognition (Graves, Mohamed & Hinton 2013), among many others, through numerical algorithms, giving rise to what we now recognise as AI.

However, a paradigm shift has emerged in recent years: rather than replacing human abilities, AI is increasingly being used to augment or even replace traditional mathematical algorithms, as it is the case in Computational Science (CS) (Tompson *et al.* 2017; Raissi, Perdikaris & Karniadakis 2019; Cranmer *et al.* 2020*a*; Lemos *et al.* 2023). This shift has profound implications, especially in the context of improving well-established numerical solvers and algorithms. However, are these same AI methods, designed to mimic human perception or language, also relevant when the task is to approximate mathematical operators? A typical illustration of this change in task nature lies in the visual metrics or simple mean square errors (MSEs) employed in the AI community, in contrast with the machine-precision accuracy and physics-based metrics usually adopted in CS. This raises the critical question: what modifications or safeguards are necessary to adapt off-the-shelf AI algorithms for physically constrained or computationally rigorous tasks? Thus, while most recent reviews on AI for computational physics (Cai *et al.* 2021; Herrmann & Kollmannsberger 2024), engineering (Zhou *et al.* 2020; Azar *et al.* 2021) or CFD (Garnier *et al.* 2021; Lino *et al.* 2023) emphasised on the various techniques and their improvements to tackle increasingly more complex situations, this work investigates this transformation from an epistemological perspective, addressing a critical question: is AI for computational science a tangible reality or merely a transient hype?

From the early ages of AI, researchers have studied the epistemological aspects of this new discipline. For instance, McCarthy (1981) considered that the epistemological part of AI, which studies what kind of tasks AI can address in principle, must be separate from the heuristic part of AI, which focuses on how, in practice, one can devise algorithms to solve those tasks. One benefit of relying on epistemological analysis is that a single solution of an epistemological problem can often admit multiple technical implementations. In other words, it would be wise to elucidate a few foundational principles required for AI in CS and CFD from an epistemological perspective, rather than enumerating the numerous technological solutions possible to address these challenges. For instance, Ganascia (2010) stresses the importance of distinguishing between the shortcomings of specific algorithms and the adequacy of the fundamental principles underlying AI. In his view, most early criticisms of 'old-fashioned AI' focused on algorithmic immaturity, without addressing whether the core principles of AI were inherently limited for these tasks. As a consequence, this present review will follow this epistemological perspective, by first identifying in § 3 the pillars of effectiveness in science through an epistemological approach, before exploring in § 4 how these fundamental aspects can actually be technically integrated into AI frameworks.

Note that some essays have already adopted such an epistemological perspective on particular aspects of AI, for example, its social inadequacy (Ganascia 2010) or the philosophical challenges when using intelligent systems for health (Sa, Carvalho & Naves 2023). However, works focusing on the epistemological foundation of AI in science, and more specifically into CS or CFD, are still rare. In that context, Duan (2023) explored the challenges to build artificial scientist. Velthoven & Marcus (2024) analysed the implications of AI in science through the philosophical ideas of Popper (1962). Their main conclusion is that AI, alone, cannot advance science as it cannot explain why certain

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

theories should be preferred over others because of its underlying inductive approach of knowledge, an idea also reported by Anderson & Abrahams (2009). Note however that Velthoven & Marcus (2024) employed systematically 'current AI' in their conclusions, opening the door to future AI systems, for example, combining both deductive and inductive approaches. This perspective is also shared by Krenn *et al.* (2022) and Kaiser *et al.* (2024), who argued that an AI could, in principle, generate revolutionary theories, but only under specific conditions. For instance, they introduced the notion of 'agent of understanding', a condition defined as the capability of an AI system to conceptualise new scientific insights and, crucially, to transfer this understanding to human researchers. While these conditions are not fulfilled nowadays, authors envisioned the possibility that future intelligent systems may indeed both generate novel theories and render them intelligible to humans. In contrast, Alvarado (2023) considered AI in science as a simple epistemic tool, that is to say, without revolutionising the way science is conducted. As a summary, the epistemology of AI for science is still recent but bourgeoning, often debating on the ability of AI to generate novel scientific discovery. Nevertheless, to my knowledge, no work investigates the epistemology associated with AI algorithms replacing existing mathematical-based algorithms and solvers in computational science, which is a different task than the only theoretical discovery. In other words, the main objective of these perspectives is to extract what foundational principles are required for AI to be useful in CS and CFD, where mathematical algorithms are currently massively employed.

To approach this question of the potential effectiveness of AI in CS and CFD, I first provide a general background on deep learning in § 2. Then, I explore in § 3 why mathematics has been so effective in science, with a specific focus on fluid mechanics, particularly in the development of mathematical-based algorithms for solving fluid equations in CFD. This analysis identifies four foundational pillars of mathematical effectiveness (PoEs), namely: (i) symmetries, which impose internal structure and coherence; (ii) scale separation, allowing distinct treatments for the different scales and their interactions; (iii) sparsity, which simplifies complexity and enhances explicability; and (iv) semantic significance, which fosters abstraction, reasoning, and interpretability. Next, I examine how these PoEs align with advancements in AI in § 4. Each aspect is analysed to determine how recent AI methods integrate these principles, particularly in the domain of CFD. This examination is supported by a review of recent applications of AI in fluid mechanics, showcasing its potential to complement, or even surpass, traditional solvers in specific contexts. Beyond these PoEs, I also examine a fifth pillar, credibility (Jebeile 2024), which is essential for translating effectiveness into the widespread adoption of AI within the fluid mechanics community. The findings suggest that AI does not inherently lack the capacity to incorporate these principles, so that they can be used as guidelines for future AI developments in CFD and, more generally, in computational science.

2. General background on deep learning

2.1. *Neuron and neural architectures*

This perspective focuses primarily on deep learning approaches in CFD. Deep learning is a branch of AI that leverages neural networks to approximate a function $f(\cdot)$ mapping an input x and an output y. This neural network, denoted f_θ, is parametrised by a user-defined set of weights $\theta = \{\theta_1, \ldots, \theta_N\}$, where N can be very large. As such, neural networks belong to the broader family of parametric models.

From the early models (Rosenblatt 1958), the core idea behind neural networks has remained: decompose the complex function f_θ into a composition of simple building

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

Architecture	Input x_n	$h_{\theta=W}$	Aggregation
MLP	Vector $x_n \in \mathbb{R}^d$	Wx_n	Over all components of x_n
CNN	Tensor $x_n \in \mathbb{R}^{C\times H\times W}$	$W \star x_n$	Local convolution, shared weights
LCNN	Tensor $x_n \in \mathbb{R}^{C\times H\times W}$	$W \star x_n$	Lipschitz constraint (e.g. $W^T W = I$)
RNN	Sequence $x_n = \{x_n^1, \dots x_n^d\}$	$W_x x_n^k + W_h h_{k-1}$	Over time: recurrence and memory
GNN	Graph (V, E), node feature x_n^v	$\sum_{u=\mathcal{N}(v)} W_{uv} x_n^u$	Over neighbours $\mathcal{N}(v)$
Transformer	Sequence $x_n \in \mathbb{R}^{T\times d}$	$Q, K, V = x_n W$	Global attention $\text{Softmax}(QK^t/\sqrt{d})V$
Fourier net.	Vector or tensor	$W \cdot FFT(x_n)$	MLP or CNN in spectral space

Table 1. Various well-known neural architectures, and their respective input format x_n and associated aggregation function h_θ parametrised by $\theta = W$ (bias is omitted for brevity). Architectures considered include the multilayer perceptron (MLP), convolutional neural network (CNN), Lipschitz-constrained neural network (LCNN), recurrent neural network (RNN) or graph neural network (GNN).

blocks called neurons. Each neuron receives an input x_n and produces an output y_n, where x_n and y_n can represent either the actual inputs or outputs of the target function $f(\cdot)$, or intermediate values, often referred to as hidden variables.

Most modern network architectures still follow the seminal principle introduced by McCulloch & Pitts (1943), inspired by biological neurons: each artificial neuron is composed of: (i) a linear parametrised aggregation function h_θ, followed by (ii) a fixed nonlinear activation function σ. This leads to the fundamental equation of a neuron:

$$y_n = \sigma(h_\theta(x_n)). \tag{2.1}$$

Despite its simplicity, (2.1) offers a flexible framework to construct a wide range of neural network architectures. The form of the aggregation function h_θ can be tailored to the input structure, as illustrated in table 1. Moreover, the choice of aggregation function directly impacts the properties of the resulting network, as further detailed in § 4.1.2. For instance, employing convolutional operations in convolutional neural networks (CNNs) allows weights sharing across the whole image, significantly reducing the number of learnable parameters (i.e. smaller W). In addition, since convolution is translational equivariant, this symmetry naturally extends to the convolutional layers of the network.

While most aggregation functions are linear, it is worth noting that attention mechanisms (Vaswany *et al.* 2017) in transformer architecture (table 1) behave differently. Although the values V are obtained through a linear transformation of the input ($V = x_n W_V$), the global attention weights depend on the inputs: they are computed as

$$W(x_n) = \text{Softmax}(QK^t/\sqrt{d}), \tag{2.2}$$

where the query is $Q = x_n W_Q$, the key is $K = x_n W_K$ and d is the dimensionality of the query/key vectors. Since both Q and K are linear in x_n, the dot product QK^t introduced a quadratic dependency of the input, resulting in a nonlinear data-dependent aggregation weights $W(x_n)$. This makes attention mechanisms fundamentally different and powerful, compared with fixed or shared-weight linear aggregations like convolutions.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

Once the aggregation and activation functions are chosen, several neurons can be interconnected to form a network. Neurons are typically organised into layers, where all neurons in a layer share the same inputs, but produce different outputs. These outputs then serve as inputs for the neurons in the next layer. Stacking multiple layers introduces additional nonlinearities into the model, thereby increasing its capacity to represent complex functions. When a network contains many layers, it is called a deep neural network. The way neurons are interconnected across layers defines the network architecture. Various connection patterns can be employed, such as skip connections or residual blocks, which often improve training stability and convergence. Additional modules, such as normalisation layers, or downsampling and upsampling operations can also be included to enhance performance or efficiency. Overall, the flexibility in defining both the neuron internal structure and the network architecture makes neural networks an extremely versatile framework, capable of addressing a wide range of tasks and data modalities.

2.2. *Training procedure: learning is optimising*

Training a neural network consists in optimising its weights θ to minimise a metric, often called loss function, denoted $\mathcal{L}(\cdot)$. In supervised learning, this process relies on a training dataset $\mathcal{D}_T = \{x^{(i)}, y^{(i)}\}_{i=1..M}$ composed of input–output pairs, where $x^{(i)}$ (respectively $y^{(i)}$) is the ith input (respectively target output) sample. The goal of training is therefore to find the optimal set of weights θ^* that minimises the loss over all samples in the dataset, which can be expressed generically as

$$\theta^* = \text{argmin}_\theta \big[\mathcal{L}\big(f_\theta(x^{(i)}), y^{(i)}, \theta\big)\big] \quad \text{for all } \{x^{(i)}, y^{(i)}\} \in \mathcal{D}_T. \tag{2.3}$$

Equation (2.3) provides a general and flexible formulation of the learning process, where the choice of the target definition y and the loss function $\mathcal{L}$ determines the type of neural network and the nature of the task it can address. Depending on these choices, one can design models for surrogate modelling, generative modelling, classification or even deep reinforcement learning (see table 2). As a simple example, a standard supervised training set-up using the mean squared error (MSE) loss is defined as

$$\mathcal{L} = \big\| f_\theta(x^{(i)}) - y^{(i)} \big\|^2. \tag{2.4}$$

In contrast, incorporating L2 regularisation (also known as weight decay) modifies the objective as

$$\mathcal{L} = \big\| f_\theta(x^{(i)}) - y^{(i)} \big\|^2 + \lambda \|\theta\|^2, \tag{2.5}$$

where λ is a user-defined hyperparameter controlling the strength of the regularisation term and $\|\cdot\|$ denotes the L2 norm. Note that classical least square method and Ridge regression correspond to (2.4)–(2.5) where the function f_θ is linear, i.e. where the activation function is $\sigma = Id$.

These surrogate models trained with an MSE loss require both the input x and the corresponding output y to compute the reconstruction error $\|y - f_\theta(x)\|^2$. When the true output is unavailable, a self-supervised strategy can still be employed by training the network to reconstruct its own input, enforcing $x \approx f_\theta(x)$. This approach becomes useful when the reconstruction is performed through a low-dimensional latent code z, which enables an efficient compression of the input representation. This task is typically achieved through an autoencoder (AE) architecture, composed of two networks: an

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

encoder f_E that compresses the input into a latent code $z = f_E(x)$ and a decoder f_D that reconstructs the original sample as $x \approx f_D(z)$. Once trained, the latent code z can be sampled (e.g. $z \sim \mathcal{N}(0, I)$), turning the model into a generative network capable of producing new data. To explicitly enforce this sampling property during training, one can constrain the latent variables to follow a prescribed distribution, typically a standard Gaussian law, by minimising the Kullback–Leibler (KL) divergence between the learned distribution $p(z)$ and the target prior $\mathcal{N}(0, I)$. Combining this regularisation term with the reconstruction loss yields the well-known variational autoencoder (VAE) formulation.

Following this idea, other techniques have been developed for enhanced generative modelling. For example, generative adversarial networks (GANs) proposed by Goodfellow *et al.* (2014) also rely on a low-dimensional latent code z to generate synthetic samples $x \approx f_G(z)$. In addition, a second network, the discriminator f_D, is trained to distinguish whether an image is real (taken from the dataset) or fake (produced by f_G). This set-up defines a minimax optimisation problem, in which the loss is simultaneously minimised to improve the generator f_G and maximised to enhance the discriminator's ability to detect fake images. Although GANs are mathematically well posed, they are notoriously difficult to train in practice and often suffer from mode collapse, a phenomenon where the generator produces limited diversity in the generated outputs. To address these challenges, diffusion-based models (Ho *et al.* 2020; Song, Meng & Ermon 2020) have recently emerged. In their forward process (F), random noise ϵ is gradually added to an initial image x_0, generating a sequence of increasingly noisy images x_t. Since the noise ϵ and the corresponding noisy image x_t are both known, the task can be formulated as a supervised learning problem, where the network is trained to predict the noise using a standard MSE loss:

$$\mathcal{L} = \|\epsilon - f_\theta(x_t, t)\|^2. \tag{2.6}$$

During the reverse process (B), the model starts from a pure Gaussian noise sample and progressively denoises it by subtracting the learned noise prediction f_θ at each step, yielding increasingly realistic reconstructions. To preserve stochasticity and allow for diversity in the generated samples, an additional noise term $\sigma_t z$, with $z \sim \mathcal{N}(0, I)$, is injected during the backward diffusion. These generative models for physics are further detailed in § 4.6.1.

Note that classification tasks also fit naturally within this general framework by predicting probability distributions rather than deterministic outputs. In such cases, the objective is to identify the most likely class c (an integer label), which is commonly reformulated as predicting the corresponding class probabilities $p(c)$. To properly train the network on these probabilistic targets, the loss function must therefore measure discrepancies between probability distributions rather than pointwise errors. This is precisely the role of the cross-entropy and KL divergence losses, which quantify the distance between the predicted probability distribution and the true one.

Beyond purely data-driven tasks such as classification or generative modelling, neural networks can also leverage prior physical knowledge to guide learning. In physics-informed neural networks (PINNs) proposed by Raissi *et al.* (2019), for instance, the solution field $u_\theta(x, t)$ is parametrised by a neural network and trained by minimising the residual of the governing equations, $\mathcal{R}(u_\theta(x, t))$. This approach effectively relaxes the need for labelled data, as the network is constrained by the underlying physics. Initial (IC) and boundary (BC) conditions can be naturally incorporated into the loss function as additional terms. While PINNs focus on learning solution fields, more recent methods aim to directly learn operators that map inputs to outputs. Such operator learning approaches, including the Fourier neural operator (FNO), DeepONet and implicit neural representation

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

(INR), are particularly useful for multi-scale or parametric problems and are discussed in detail in § 4.2.3.

Finally, deep reinforcement learning (DRL), often used for optimal control of fluid flows (Verma, Novati & Koumoutsakos 2018; Rabault *et al.* 2019; Berger *et al.* 2024; Koumoutsakos 2024), can also be framed within this general neural network perspective. In DRL, a network is trained to approximate either the optimal policy $\pi_\theta(a|s)$ or the action-value function $Q_\theta(s, a)$, as a function of the system state s and action a. The main challenge is that the true target is not known *a priori*, since it would require the knowledge of the optimal control law. Consequently, the target must be iteratively estimated during training, yielding a loss function similar to MSE, but with self-consistent updated targets instead of fixed labels. This formulation naturally fits within the general neural network framework, although it typically results in more unstable training dynamics compared with standard supervised learning.

Beyond the task-specific formulations summarised in table 2, recent advances have introduced a new class of broadly trained models known as foundation models (Bommasani *et al.* 2022). Typical examples are large language models (LLM) such as the well-known ChatGPT (Brown *et al.* 2020). Instead of optimising a network for a single objective, they are rather pretrained on massive heterogeneous datasets using self-supervised objectives, with the goal of learning general-purpose representations of data. After this large-scale pretraining, the resulting representations can be adapted to a wide variety of downstream tasks (regression, classification, generation or control), often with minimal additional training.

In conclusion, neural networks offer a flexible and generic framework capable of tackling a wide range of complex tasks. The natural next question is how effective this framework truly is, particularly when compared with the well-established and highly optimised mathematical algorithms commonly used in computational physics, and especially in CFD. This comparison will allow the identification of key principles from classical methods that could guide neural networks to achieve comparable, or even superior, performance.

3. Effectiveness of mathematics throughout the history of science

Since Galileo Galilei, physicians intensively employ mathematics to describe natural phenomena, including gravitation, electro-magnetic fields as well as fluid mechanics among many others, exhibiting remarkable precision and generalisation capabilities. According to Feynman, ‘If you want to learn about nature, to appreciate nature, it is necessary to understand the language that she speaks in’. This statement is in line with the Galilei’s conjecture that the language of nature is mathematics. While this viewpoint implies the straightforward effectiveness of mathematics, it relies solely on Galileo’s conjecture, which cannot be demonstrated or rejected, yielding endless circular debates with no resolution in sight. Thus, let us consider another perspective, where mathematics in physics is useful because we have no choice. For instance, Kepler posed the question ‘What else can the human mind hold besides numbers and magnitudes?’. In other words, the real world is too complex and mathematics is our sole means to make this inherent complexity of nature intelligible to human understanding. Let us consider the situation where only the mathematical properties of the physical world are accessible to humans. Here, mathematics is not considered as the language of nature, but rather a tool for humans to apprehend its complexity. In that situation, physics would be effective only because its use is limited to mathematisable questions. This is evident in fields like medicine, where very few equations exist to describe the fundamental interactions between living

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

Task	Output / Learned function	Loss / Objective $\mathcal{L}$
	Surrogate models	
Surrogate model	$y = f_\theta(x)$	MSE
Sparse surrogate model	$y = f_\theta(x)$	MSE $+ \lambda\|\theta\|^2$
Autoencoder (AE)	$z = f_E(x)$ and $f_D(z) \approx x$	MSE
Model for classification	$y = p(c) = f_\theta(x)$, with class $c \in \{1, \dots, P\}$	Cross-entropy
PINN	$u_\theta(x, t)$ such that $\mathcal{R}(u_\theta) = 0$	$\|\mathcal{R}(u_\theta)\|^2$ + IC and BC
Operator learning	$u(x) = \mathcal{F}_\theta(g)(x)$, where $\mathcal{F}_\theta : g \mapsto u$	MSE over sampled (g, x)
	Dynamical models	
Equation discovery	$\dot{x} = \mathcal{D}(x)\theta$ with dictionnary $\mathcal{D}$	MSE $+ \lambda\|\theta\|_0^2$
Autoregressive network	$x_{t+1} = f_\theta(x_t)$	MSE + Long term loss
Neural ODE	$\dot{x}(t) = f_\theta(x(t), t)$ + ODE solver	MSE on x and/or $\dot{x}$
	Generative models	
Variational AE	$z = f_\mu(x) + f_\sigma(x)\epsilon$, where $\epsilon \sim \mathcal{N}(0, 1)$ and $f_D(z) \approx x$	MSE + KL divergence
GAN (minimax game)	$x \approx f_G(z)$ and $f_D(x,\ f_G(z)) = 0/1$	$min_G max_D \mathbb{E}_x[log(f_D)]$ $+\mathbb{E}_z[log(1 - f_D(f_G(z)))]$
Diffusion model	(F): $x_t = \sqrt{\alpha}x_0 + \sqrt{1-\alpha}\epsilon$	Denoising score
(e.g. DDPM, DDIM)	(B): $x_{t-1} = \left(\frac{1}{\sqrt{\alpha}}\right)\left(x_t - \left(\frac{1-\alpha}{\sqrt{1-\alpha}}\right) f_\theta\right) + \sigma_t z$ where $\epsilon, z \sim \mathcal{N}(0, I)$	(e.g. MSE: $\|\epsilon - f_\theta(x_t)\|^2$)
	Deep reinforcement models	
Deep RL	$\pi_\theta(a\|s)$ or $Q_\theta(s, a)$ for action a and state s	Value loss, TD error

Table 2. Overview of various neural network tasks with their associated learned function and training objectives.

cells or biological systems, making it less amenable to mathematical treatments. Note that mathematics are still used for specific problems, like understanding the microbial population dynamics (Quedeville 2020) or the DNA replication and repair using knot theory (Robic & Jungck 2011), among others, including the pioneer works to create mathematical models of biological neurons, yielding the emergence of AI (McCulloch & Pitts 1943). Interestingly, this limitation had led to the early widespread adoption of AI in medical research, such as cancer detection (Tekkesin 2019) and protein discovery (Madani *et al.* 2023), since no other efficient algorithm was already successful for such tasks. Nevertheless, does it imply that AI cannot play a crucial role in disciplines already massively based on mathematics, as it is the case for fluid mechanics and CFD, and more

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

generally in computational physics? How can mathematics and AI complement each other in such domains? Can the fundamental properties of mathematics be transposed into AI to enhance both its effectiveness and adoption within the scientific community? In my opinion, these questions are rarely addressed and, therefore, this section is dedicated to explore and to answer them, at least partially.

3.1. *Definitions of effectiveness*

Before diving into the epistemology of mathematics, the notion of the effectiveness of a scientific method should be defined and clarified. According to Wigner (1960), the effectiveness of mathematics in natural science is a miracle, out of reach of the human understanding. Nevertheless, he defined three levels of effectiveness in physics.

First, the prediction capability. As a main objective throughout history, scientific models and theories have been developed to predict behaviours of physical phenomena. In that view, a naïve definition of effectiveness is the capability of a model to predict accurately unseen behaviours, i.e. to generalise to new conditions, at least in a pre-defined range of operating points. On one hand, it is worth noting that prediction effectiveness should not be confused with the sole accuracy, but should be also related to its inherent cost to provide such a prediction. For instance, is Reynolds-averaged Navier–Stokes more or less effective than large eddy simulation (LES) or even DNS in the context of CFD? Similarly, is a linear regression model less effective than a deep neural network? Consequently, a proper metric defining the trade-off between accuracy and cost (or any measure of the complexity of a model) has to be proposed when such effectiveness is invoked. An alternative is to set an accuracy threshold and evaluate the cost of the method to obtain this user-defined accuracy (Ajuria-Illaramendi, Bauerheim & Cuenot 2022). On the other hand, unseen behaviour is usually not clearly defined. This is due to the difficulty to properly construct a metric to assess how far a new behaviour is, compared with previous seen phenomena. For instance, this difficulty typically arises in very high dimension, as usually encountered with deep learning (Xia *et al.* 2015), so that identifying 'unseen' data to evaluate generalisation becomes illusive. This question is attracting attention recently, in particular by finding methods to identify these out-of-distribution data (Durasov *et al.* 2024; Yang *et al.* 2024).

Second, the capability to provide explainable structures. Mathematics in physics is not limited to reproduce results provided by experiments, but allows us to understand the underlying mechanisms and structures encompassing, and thus explaining, multiple similar situations. It constitutes a conceptual generalisation. As an illustration, the general relativity proposed by Einstein (1920) goes beyond its simple validation, such as the recent tests on the weak equivalence principle provided by the Microscope mission (Touboul 2022). Indeed, this theory offered to see gravitation as the curvature of the four-dimensional space–time, rather than considering it as a mere force. This epistemological rupture allowed the explanation of the previously insoluble anomaly of Mercury's orbit, but also the description of new phenomena as surprising as gravitational lenses, black holes or gravitational waves. As a consequence, this theory goes beyond predictions, bringing also explanation, which echoes the 'Predicting is not explaining' written by Thom, Tsatsanis & Lisker (2016). In addition, he underlined the need of unification in physics, which stipulates that explaining means reducing the large diversity of phenomena to a few general principles. It is worth noting that in this viewpoint, CFD is effective in predicting flows, but not directly to explain their behaviour: to do so, post-processing and analysis are always required. For example, while compressible CFD allows the simulation of aero-acoustic phenomena, determining the type and localisation of acoustic sources in a simulated flow still constitutes a challenge. This requires state-of-the-art post-processing,

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

like Helmholtz's or Doak's decompositions, to separate acoustic perturbations from turbulent and hydrodynamic ones (D'Aniello *et al.* 2023). Similarly, finding the correlation between flow structures, like vortices, and quantities of interest, like lift, also necessitate the development of novel post-processing, like the force partitioning method (Menon & Mittal 2021), to extract meaningful information from the CFD simulations.

Third, the conceptual generative. Effectiveness can also be associated with another capability, the one called 'generative' by Wigner (1960). This notion translates the ability of a theory, not to predict or to explain, but to fertilise novel ideas and concepts, often based on new representations. Bachelard (1936) emphasised this notion by stating that 'mathematical information gives us more than reality; it gives us the plan of the possible; it goes beyond the actual experience of coherence'. In essence, the objective of mathematics in physics extends beyond merely providing theories consistent with experimental measurements. Rather, it provides a systematic framework that, starting from a small set of foundational principles, generates the full range of possible theories consistent with those principles. This approach allows physicists to generalise existing concepts, but also to envision novel ones, even in the absence of empirical validation or refutation (Maronne & Patras 2022): 'To think about reality is to construct it mathematically'. Consider two examples from the history of scientific discoveries. On one hand, in 1846, Le Verrier predicted the existence of Neptune using only mathematical calculations. Its presence was soon confirmed by Galle's observations. This was an impressive achievement, but this was just one more planet among those already known. On the other hand, the theoretical predictions of the Higgs boson introduced an entirely new kind of entity into physics, while providing a conceptual shift where mass is not an intrinsic property of particles. This idea was experimentally confirmed only 48 years later at the Large Hadron Collider (LHC), revealing that conceptual generative can appear prior to any prediction capabilities of a theory. It is worth noting that mathematical theories can exhibit conceptual generative without necessarily producing significant new explanations or predictions, as seen in string theory. In that case, the new physical concept serves as a source of new ideas for pure mathematics. Such theories with conceptual generative but not yet experimentally verified are often denoted as 'tentative theories' (Penrose 1990). This illustrates that no inclusion relationship exists between these three levels of effectiveness: predictive, explanatory and conceptual generative.

Nevertheless, AI for CFD remains an emerging approach and to be considered truly effective, it must consistently prove its predictive power by surpassing traditional mathematical tools. This is a prerequisite for broader adoption within the research and engineering communities. However, while AI has demonstrated superior performance in specific tasks and configurations, a major challenge remains: extending its applicability to a wider range of problems with minimal or no retraining. Interestingly, the rapid advancements in AI to achieve this objective have sparked a wave of innovative ideas, particularly in the modelling and understanding of these algorithms. In fluid mechanics, novel concepts bridging AI and CFD have emerged, showcasing its potential for conceptual generative effectiveness. Examples include physics-informed neural networks (PINNs) introduced by Raissi *et al.* (2019), which approximate partial differential equation (PDE) solutions with continuous functions without relying on gradient approximations, and hybrid differentiable solvers (Um *et al.* 2020), which seamlessly integrate AI with classical numerical methods. Thus, the next crucial challenge for AI lies in explanatory effectiveness. First, AI models and training strategies must be interpretable, providing clear guidelines for practical implementation and facilitating adoption in research and engineering. Second, beyond just predictive accuracy, AI systems should be capable of understanding and explaining the underlying physics, similar to the 'agent

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

of understanding' proposed by Krenn *et al.* (2022), a challenge rarely addressed in the literature. While AI has shown promise in identifying complex nonlinear physical phenomena, its ability to provide physical insights remains limited, in particular on unknown and complex problems. This lack of interpretability represents a key area for future improvement, and I believe this is an essential step for AI to evolve into a reliable tool for scientific discovery and engineering applications.

That being said, note that Wigner (1960) also insists on the fact that a new formalism, tool or concept is rarely directly applicable in the domain for which it was thought. As a consequence, evaluating the effectiveness of a theory *a priori* has no significant meaning or, at the very least, necessitates a careful consideration. This is why the following sections do not investigate the effectiveness of mathematics and AI based on their respective current results, but rather on their epistemological potential. Let us begin by delving into the origins of mathematical effectiveness to extract its foundational principles.

3.2. *A reasonable explanation of the effectiveness of mathematics*

3.2.1. *Symmetries and invariants*

Lambert (1997) has examined the possible reasons behind the effectiveness in mathematics in his book 'Is the efficiency of mathematics unreasonable?', inspiring the title of this very subsection. Within this work, he explores various philosophical theories, such as Pythagoreanism, Platonism, empiricism etc. intended to explain the effectiveness of mathematics and, in general, to every scientific methods. Among them, he mentioned formalism, which stipulates that mathematics is effective because it relies on symbols and rules to manipulate them. Nevertheless, Lambert (1997) argues that this point of view can explain the effectiveness of mathematics only as a kind of miracle, as it is difficult to see how a 'game' governed solely by constraints of internal coherence could correspond to a reality external to thought, such as physical phenomena. However, this theory gave birth to relationism, which considers mathematics as the formalised representation of all possible conceptual relationships, including numerical ones. For instance, according to Bohr (1958), 'we shall not consider pure mathematics as a separate branch of knowledge, but rather as a refinement of general language, supplementing it with appropriate tools to represent relations for which ordinary verbal expression is imprecise or cumbersome'. Here, mathematics is not just a language with symbols and rules, but serves as a means of formally translating and elucidating the relationships between concepts. This might explain why physics is well mathematised, since most results can be expressed through structured relationships between the fundamental quantities of the system, structures that mathematics can faithfully capture and elaborate. According to Vandamme (1976), 'The knowledge in general, and even more so the scientific knowledge, is characterised by its structures'. Nevertheless, it is worth noting that not all relationships are useful. Corry (1996) notes that mathematicians are not interested in cataloguing all possible structures, but rather to investigate ones that possess rich invariants. This is consistent with physics, where an 'element of reality' cannot be described properly without recourse to invariants preserved under the symmetries of the system. This need of invariance is motivated by our partial observation of reality that requires invariants to conclude without exhaustive observations, which is often impractical or cumbersome. In other words, reality in physics must be robust to a change of point of view. This is exactly what research in physics is doing, but more generally how humans and animals perceive the surrounding world. For instance, Florack *et al.* (1994) and Lindeberg (2013) have demonstrated that human vision and cognitive functions rely on some invariants and symmetries: 'It is apparently the case that, at least up to some approximation, many visual tasks can be solved on the

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

basis of such invariants only. This observation should affect the way computer-vision and image-analysis tasks are handled as well' (Florack *et al.* 1994).

Modern science, and mathematics, now heavily rely on the study of symmetry, as illustrated by Anderson (1972) who noted that 'it is only slightly overstating the case to say that physics is the study of symmetry'. Early on, Curie (1984) proposed his principle of symmetry, based on ideas from F. Neumann and B. Minnigerode, where 'the symmetries of the causes are to be found in the effects'. This highlights the profound relevance of symmetries to understand the surrounding world. Interestingly, symmetries in physics are intricately related to invariants, i.e. conserved quantities, also known as first integrals, as elegantly elucidated by Noether's theorem (Noether 1918). As an example, rotational symmetry yields the conservation of angular momentum, and invariance under a translation in time leads to the conservation of energy. Moreover, symmetries and invariances are much more intricate and influential in modern physics.

In the realm of fluid mechanics, the rigorous investigation of symmetry and invariance has also been studied for a long time. For instance, Golubitsky & Stewart (1986) studied the specific symmetries acting on the stability of Taylor–Couette flows. They demonstrate that symmetries inherent in bifurcating systems have a strong influence on their behaviour, which motivates their study. The effect of symmetries on flow instability has been recently revisited by Sierra-Ausin, Fabre & Knobloch (2024), focusing on the $O(2)$ symmetry applied to wake instability. Similarly, combustion instabilities in annular combustion chambers are governed by the azimuthal symmetry. When this symmetry is broken because of different flame types (Bauerheim *et al.* 2014*b*), azimuthal mean flow (Bauerheim *et al.* 2014*a*) or stochasticity (Bauerheim *et al.* 2016*a*), modal splitting occurs which affects the whole stability of the system. Recent works in thermoacoustics still involves this notion of symmetry breaking (Aguilar *et al.* 2021; Faure-Beaulieu *et al.* 2021). Symmetries have also been exploited in fluid mechanics to build turbulence models. Indeed, Grebenev & Oberlack (2007) and Fakhar *et al.* (2009) studied the symmetry of the incompressible Navier–Stokes equations (Cantwell 1978). Based on these symmetries, Razafindralandy, Hamdouni & Oberlack (2007) and later Sayed *et al.* (2010) derived new turbulence modelling which preserve those symmetries. They also analysed the existing turbulence models available at that time, including the Smagorinsky model (Smagorinsky 1963), showing that only its dynamic version (Germano *et al.* 1991) allows symmetry preservation.

These examples reveal the uncanny effectiveness of mathematics because of symmetry and invariance, highlighting its capacity to shape our understanding of reality while existing purely within the realm of abstract ideas. This fundamental concept of symmetries and invariance is also embedded in the very construction of modern AI methods for classification and regression tasks (Gens & Domingos 2014; Weiler *et al.* 2018; Weiler & Cesa 2019; Wang, Walters & Yu 2021). Typical examples are CNN built upon the translation equivariance principle or graph networks based on permutation invariance. Recently, the community devoted to AI for science has been exploring the necessity of incorporating symmetries or structures directly into neural network architectures (Önder & Liu 2023; Otto *et al.* 2023), not only as a design principle, but also to align with the intrinsic demands of science, which fundamentally relies on these principles. Similarly to the Erlangen programme proposed by F. Klein in 1972 to unify geometry as the study of invariants or symmetries using the language of group theory, a recent objective in AI is to unify the various network architectures into a general framework, known as geometric deep learning, based also on symmetries or invariants (Bronstein *et al.* 2021). As a consequence, symmetry and invariants constitute the first pillar to make AI effective in CFD, as further discussed technically in § 4.1.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

3.2.2. *Scale separation*

Furthermore, while investigating invariants for image recognition, Florack *et al.* (1992) also emphasised the scale dependence of these invariants. This notion of scale is crucial in physics, where the physical laws governing complex systems exhibit a sort of separation between the various levels of structures associated with spatial and temporal scales. For instance, while the classical Euclidean space is intuitive at spatial scales we encounter in our everyday life, this framework fails to explain phenomena at very low (quantum mechanics) and very large (gravitation) scales. Einstein (1920) challenged this view by introducing the concept of curved space–time, effective at large scales, but still failing to explain the transition between the continuous world we experience and the quantised energy realm of quantum mechanics. Similarly in fluid mechanics, working at the macroscopic (Navier–Stokes equations) or mesoscopic (lattice Boltzmann method) scales does not require a detailed knowledge of the fine microscopic scales, such as the interactions between the fundamental particles. This scale separation, when available, allows dedicated physical analyses and methodologies to be performed, introducing its own abstraction. For instance, the notion of pressure at the macroscopic scale in a fluid is intuitive, while being much more complex, or even meaningless, at much lower scales.

These considerations are essential not only in physics, but also in the development of AI, as well as across a broad spectrum of disciplines, including sociology, history etc. This idea aligns closely with the epistemological perspective emphasised by Bachelard (1938), who considered that the main challenge of science is to push humans beyond their intuitive sense of scale. Additionally, Simondon (1970) named scale rupture this transition between ordinary scales, where familiar schemes like particles or Euclidean space remain valid, and the extreme scales studied by quantum mechanics and astrophysics, where these schemes inherited from ordinary perception are no longer adequate. It suggests that the effectiveness of mathematics in physics derives from its ability to conceptualise and study phenomena occurring at very different scales compared with those the researcher, as a human being, is used to in a daily-life.

This principle is fundamental in CFD, where applications span an enormous range of scales, from microscopic flows around biological cells (O'Connor *et al.* 2016) to astrophysical phenomena, such as supernova-induced turbulence and other large-scale flows in the universe (Müller 2020). Early on, engineers and researchers recognised the necessity of classifying these flows based on characteristic non-dimensional parameters, such as the Reynolds or Mach numbers, among others. It is well established that flow behaviour changes drastically as these parameters vary, leading to the emergence of complex phenomena like shock waves or turbulence. These flow regimes can be highly localised, coexist within the same system, interact nonlinearly or even exhibit fractal characteristics (Meneveau & Sreenivasan 1987), making them inherently multi-scale and challenging to capture and analyse.

Consequently, scale separation is central to many advancements in fluid mechanics. For instance, separation between the slowly varying large-scale perturbations from the rapidly varying small-scale of the flow is essential to derive a simplified effective description of the dynamics at large scales, and justify the concept of eddy viscosity (Dubrulle & Frisch 1991). Similar multiple scale approaches are employed by Ovenden (2005) to model sound transmission through a slowly varying duct of arbitrary cross-section, by Faure-Beaulieu & Noiray (2022) to extract the dynamics of the slow-flow variables in an annular thermo-acoustic system, and by Magri *et al.* (2016) to formulate a set of equations in the low-Mach number limit for linear stability analysis.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

Moreover, this multi-scale character of fluid flows lies at the core of many challenges in CFD. To address these, high-order numerical schemes (Bassi & Rebay 1997; Nicoud 2004) and lattice Boltzmann methods (LBMs) (Marié *et al.* 2009; Coratger *et al.* 2021) have been developed, specifically to preserve the propagation of fine-scale turbulent structures over long distances and to efficiently capture interactions across a wide range of scales. Similarly, identifying regions with extreme length scales, such as thin shock waves, flames or boundary layers, has led to the development of specialised numerical treatments (Colin *et al.* 2000; Nicoud *et al.* 2001). The critical role of scales in CFD is further highlighted by the widespread use of post-processing techniques, such as Fourier and wavelet transforms (Grizzi & Camussi 2012), proper orthogonal decomposition (POD), and other spectral analysis tools. These methods help to extract spatial and temporal scales where energy is generated or dissipated, providing deeper insight into the underlying flow physics.

This inherent multi-scale nature of fluid flows presents a significant challenge for AI, which is not initially specifically designed to handle such drastic scale separation with order of magnitudes. A clear example lies in denoising filters (Pratt 1972), used in some generative models, which often treat high-frequency components as noise and discard them (e.g. Wiener's filter). In CFD, however, small-scale structures play a crucial role in accurately capturing overall flow behaviour and cannot simply be removed. Recognising this limitation, recent research has increasingly focused on equipping deep networks with multi-scale (Nabian 2014; Mathieu, Couprie & LeCun 2016) and multi-resolution (Catalani *et al.* 2024; Cheng *et al.* 2024) capabilities, particularly for flow applications. These advancements open new avenues for improving deep learning methods in fluid mechanics by enforcing specialised treatments at different scales, as further discussed in § 4.2.

3.2.3. *Sparsity*

Pertaining directly to the second form of effectiveness as defined by Wigner (1960), i.e. the capability to provide explainable structures, mathematics is based on concise equations (sparsity), facilitating explicability: sparsity requires abstraction, generating a semantics for physics. Semantics refers to the branch of linguistics and logic concerned with meaning. Sparsity and parsimony, as suggested by William of Occam, are of true interest since all models are wrong: '*Pluralitas non est ponenda sine necessitate*' (Plurality is not to be posited without necessity). Among models with equivalent accuracy, the most effective are those that are sparse. For instance, the compactness of the expressions describing the Navier–Stokes equations is truly remarkable when considering the numerous physical fluid phenomena they govern: these applications range from the dynamics of blood circulation within the human heart (Chnafa *et al.* 2015) to the airflow patterns around an aircraft, and even extend to the turbulent storms on Jupiter or the cataclysmic events of supernovae (Müller 2020). How feasible are such compact expressions encapsulating so much physics? Occam's razor suggests that sparsity is at the root of explicability, but nothing yet allows us to justify that these expressions will be so general and so effective in physics. This fundamental concept of sparsity is central to many approaches in recent AI for physics, such as discovering physical laws by neural networks. This is often achieved through sparsity-promoting algorithms (Brunton *et al.* 2016*a*,*b*; Silva *et al.* 2020) allowing to focus the explainability of the model in a few terms, and hence be more general by construction.

This explainability induced by sparsity can be illustrated with one remarkable example: the evolution of Maxwell's equations. Initially, Maxwell (1861) formulated a set of 20 equations to describe the behaviour of electric and magnetic fields. The quest for sparsity

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

and elegance in Maxwell's equations culminated in the realisation that the equations could be unified and simplified through the introduction of vector and tensor calculus. This reformulation has lead to a compact set of the four well-known equations, which simplifies interpretability through the divergence and curl operators, and manipulation such as deriving Poynting's theorem for the energy conservation. However, electric ($\boldsymbol{E}$) and magnetic ($\boldsymbol{B}$) fields still appear as separate entities, while in fact they can be unified into the same compact theory. This paradigm shift occurred when the equations were again reformulated, but now using the language of differential geometry. This leads to only two compact equations, where the electric and magnetic fields appear as components of a single electromagnetic tensor ($F_{\mu\nu}$), where $F_{0i} = -E_i/c$ and $F_{ij} = -\epsilon_{ijk}B_k$. This unification of electric and magnetic fields is only possible because space and time are themselves combined into a single four-dimensional space–time, so that this sparse formalism naturally incorporates the principles of relativity. Beyond its compactness and elegance, this tensorial representation makes the fundamental Lorentz invariants of the electromagnetic field emerging directly from simple tensor contractions, highlighting the deep geometric structure underlying electromagnetism.

Similar illustrations of sparsity also arise in fluid dynamics, for instance, from the Boltzmann equation. Instead of solving directly the full set of nonlinear Navier–Stokes equations for mass, momentum and energy conservation, LBM reformulates fluid dynamics in terms of a single mesoscopic transport equation for the particle distribution function $F(x, t, v)$. This sparse formulation makes conservation laws not imposed explicitly, but emerging solely from the moments of this distribution function. This is governed by the symmetries of the velocity space that ensures that only a handful of velocity moments are conserved, while the rest decay. This explains why mass, momentum and energy are the universal conservation laws in fluid flows. In addition, while the velocity space is continuous, its discretisation into a small set of representative velocities (e.g. D2Q9, D3Q19) is sufficient to preserve the essential hydrodynamic moments. This reduction transforms the complexity of the full PDE system into a simple stream-and-collide process, which is computationally effective. In this way, sparsity provides: (i) a compact representation; (ii) a clear physical interpretation; and (iii) a computational effective method.

Another example of sparsity is provided by turbulence modelling. While DNS resolves all turbulent scales, a task that rapidly becomes intractable at high Reynolds numbers, RANS circumvents this by replacing the fluctuating motions with the Reynolds stress tensor, which contains only six independent unknowns. On one hand, Reynolds stress models (RSMs) solve transport equations for each component, offering high fidelity but little interpretability. On the other hand, eddy-viscosity models collapse the entire tensor to a single scalar coefficient, capturing turbulence as an effective viscosity. Despite its simplicity, it provides valuable physical insight by revealing turbulence's essential role as a mechanism for momentum diffusion and enhanced mixing.

As a consequence, much like the compact form of Maxwell's equations revealed the underlying structure of electromagnetism, sparse formulations and models of fluid motion provide deeper physical insight by exposing the fundamental role of key quantities in organising flow behaviour. In other words, sparsity promotes explicability. While sparsity often requires new assumptions, sometimes at the expense of predictive accuracy, these models typically retain a remarkable degree of expressive power despite being very low-parametric. However, what about the whole CFD pipeline? In practice, CFD employs high-dimensional inputs such as meshes, yielding algorithms that are not sparse in a formal sense, but nevertheless remain interpretable. I argue this interpretability arises in part because CFD is not monolithic: it is built from an aggregation of models and numerical

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

schemes, each designed for specific objectives (e.g. turbulence closure, discretisation, boundary conditions). Such modularity allows a 'divide and conquer' strategy when diagnosing errors, debugging or analysing sensitivities of particular components of the modelling chain. This notion is referred as decomposability by Lipton (2016), where 'each part of the model [. . .] admits an intuitive explanation'. In contrast, modern deep learning models tend to be monolithic, highly over-parametrised and opaque: predictive success is achieved only through massive parametrisation, but with interpretability being sacrificed (Doshi-Velez & Kim 2017). It highlights that the pursuit of sparsity in AI may involve trade-offs distinct from those in traditional mathematical tools. This also suggests that, while sparsity is indeed a critical pillar of effectiveness in scientific modelling, it may not be sufficient on its own: not all sparse methods are necessarily either effective or explicable, as further discussed in the AI context in § 4.3.

3.2.4. *Semantic significance*

In science, explicability is crucial, as it constitutes a necessary condition for constructive critical thinking. However, critical thinking can only be exercised in relation to an informed opinion, exhibiting the path and reasoning behind it. This process is often slower and more laborious than the rapid predictions of AI models. Yet, would we be willing to adhere to a result or a decision produced by an unexplainable AI, either in biology and medicine, but also in engineering? Maybe if the gap between AI and humans is substantial. This difference between slow reasoning and ultrafast predictions is characteristic of the distinction between conscious reasoning versus intuition, where the latter is 'the ability to acquire knowledge, without recourse to conscious reasoning or needing an explanation' (Patterson & Eggleston 2017).

A similar distinction is often invoked in artificial intelligence: on one hand, symbolic learning methods attempt to mimic conscious reasoning by manipulating explicit rules, logical relations or symbolic representations; on the other hand, statistical learning methods, such as neural networks, rely on detecting patterns in large datasets to provide rapid predictions without explicit reasoning. Mathematics aligns closely with the symbolic side, where logic and reasoning are the angular stones. In the realm of mathematics, we want to reason in an argumentative way to uncover truths. Therefore, mathematicians not only construct reality, but the mathematical notions created directly speak to reason (Bachelard 1934). This forces mathematics, and thus physics, to be more than just a language with syntax, but also to possess semantics, where the structures and the relationships between mathematical objects are essential: 'The knowledge, indeed, does not consist in a mere collection of data. On the contrary, it is well known that something can be considered as a data precisely thanks to its relationships with the structure, more or less abstract, of knowledge' (Vandamme 1976). In this context, semantics allows the reduction of information to manageable quantities with underlying structure and relationships, presented in a format or symbols (i.e. using the syntax) which facilitates their evaluations. This is precisely what mathematics accomplishes.

A typical example of semantic significance emerges in the construction of the various mathematical sets of real numbers ($\mathbb{R}$), complex numbers ($\mathbb{C}$) and quaternions ($\mathbb{H}$). Consider the case of a real squared number being negative, which has no meaning in $\mathbb{R}$. By expanding our realm of exploration to include not only $\mathbb{R}$, but also a new set $\mathbb{C}$, we invent a new concept ('i', the pure imaginary number), which restores the validity of the initial proposition. However, what does this extension bring in terms of semantics? Is the transition from $\mathbb{R}$ to $\mathbb{C}$ merely an artefact to enforce the validity of otherwise meaningless expressions or does $\mathbb{C}$ introduce a new layer of meaning? Additionally,

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

critically, does this meaningful transition only exist as an abstract idea or does it translate into tangible objects when analysing physical phenomena? To answer these questions, we can observed that expanding these sets inevitably alters their underlying syntactic rules. For instance, $\mathbb{R}$ is a complete ordered field, whereas in $\mathbb{C}$, we lose ordering and in $\mathbb{H}$, commutativity. These syntactic rules, i.e. constraints governing how mathematical operations are performed, are central to the structure of these sets (Matsumoto & Nakai 2023), which yields profound semantics much broader than their initial simple definitions. Indeed, each set provides a model for mathematical and physical systems where analogous rules are present or absent. Rotations offer a relevant example: while rotations in two dimensions are commutative, those in three dimensions are not. Consequently, unitary complex numbers can represent two-dimensional (2-D) rotations, and unitary quaternions rotations in three dimensions, since possessing the same structure. This connection between these sets and physical rotations can be formalised through the more general Lie theory, where complex numbers naturally encode planar rotations, while quaternions provide the algebraic structure underlying three-dimensional (3-D) rotations. In other words, syntactic rules and semantics are intertwined. This interconnection stems from the ability of mathematics to provide structures and relationships among diverse objects, either abstract entities as in pure mathematics (e.g. set of numbers) or tangible entities as in physics (e.g. rotations). These mathematical structures serve as a bridge between abstract ideas and real-world physical phenomena, thereby generating profound semantics to mathematics.

As a result, semantics extends beyond the mere need for explainability. Incorporating semantics into a model introduces an inherent structure and sparsity, which in turn reduces computational costs, minimises data requirements, enhances generalisation and fosters greater acceptance within the scientific community. While symbolic AI methods naturally embed such structured reasoning, the current trend remains largely driven by statistical learning approaches, especially deep neural networks. This divergence raises fundamental questions about the future direction of AI in CFD. Should we continue relying solely on data-driven techniques or should we advocate for a stronger integration of symbolic AI? Given that symbolic methods also face significant challenges in tackling real-world problems, could hybrid approaches mixing symbolic AI and deep networks provide a viable alternative for CFD? Addressing these questions requires a deeper reflection on the practical means of embedding semantic structures within modern deep learning frameworks, further discussed in § 4.4.

3.2.5. *Four pillars of effectiveness (PoEs)*

As a summary, the effectiveness of mathematics in physics can be attributed to four main aspects, hereafter named the ‘four pillars of effectiveness’ (PoEs): (i) symmetries providing internal structures to equations; (ii) scales separation, allowing a specific treatment of the physics for each scale and their interactions; (iii) sparsity, which promotes explicability; and finally (iv) semantic significance, which gives mathematics not a purely grammatical structure, but also a meaning, facilitating abstraction and reasoning. Those aspects will be further analysed in the specific context of the development of artificial methods for CFD, which also applies more generally to computational physics.

Yet, as claimed by Jebeile (2024), the actual usability of a model is not determined solely by its accuracy or effectiveness, but also depends on its credibility, which is its ability to earn trust and foster adoption within a community. In mathematics, credibility is inherently tied to rigour and logical proof, but in AI, credibility is contingent: it depends on empirical validation, the reliability and trust in data, and the interpretability of models.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

For example, Levins (1966) argued that credibility can emerge from the 'robustness across models', where consistent conclusions drawn from an ensemble of different models provide confidence to shared results, an idea closely connected with epistemic uncertainties. Furthermore, Simon (1962, 1969) emphasised near-decomposability and hierarchical organisation in the study of complex systems, arguing that modular structures enhance both tractability and interpretability. This insight has more recently been explored by Jebeile (2024), who mentioned that decomposability plays a key role in establishing the credibility of scientific models. This suggests that beyond the proposed PoEs, AI for CFD may also require explicit safeguards of credibility, not as a condition for effectiveness itself, but as a prerequisite for its widespread use. This fifth pillar will be further discussed in § 4.5.

4. Routes towards effective AI in CFD

Artificial Intelligence has become a transformative tool in CFD, introducing innovative strategies to address the complex, multi-scale, nonlinear challenges inherent in fluid flows. From cutting-edge neural network architectures to hybrid AI–physics models, AI demonstrates the potential to complement, or even surpass, traditional numerical methods across various applications. This comprehensive review explores key research directions at the intersection of AI and CFD, focusing on the unique challenges posed by fluid dynamics and how AI might overcome them. In particular, it investigates whether AI can implement the four pillars of mathematical effectiveness outlined in § 3.2, which are foundational to the success of mathematics in physics. These principles may provide a roadmap for endowing AI with the qualities that make mathematics so effective in modelling physical systems, in addition to a fifth pillar empowering credibility to these AI models, facilitating their widespread use.

In this section, I explore several pathways towards the development of effective AI for physics, each deeply rooted in the foundational pillars established in § 3.2. First, I review in § 4.1 how the various architectures of deep networks incorporate the first PoE, namely symmetry and invariance, and discuss how their internal structures can be constrained to reflect physical conservation laws. Second, I analyse in § 4.2 how scale separation (PoE-ii) is integrated through the design of multi-scale networks (§ 4.2.3), which optimise both computational and representational efficiency, as well as through the use of multi-fidelity data (§ 4.2.4), which enhance accuracy and robustness by combining sources from heterogeneous scales. Third, I argue in § 4.3 that sparsity (PoE-iii) plays a critical role in AI approaches, particularly for equation discovery and inverse problems. Fourth, I investigate in § 4.4 how recent approaches explicitly embed semantics into AI systems. Beyond these four pillars, I also address in § 4.5 the issue of credibility, understood as a prerequisite for the widespread adoption of AI in the fluid community. This involves the development of robust methods capable of quantifying uncertainty, allowing a first introduction of doubts into learning algorithms. Finally, while generative and foundation models appear at first to diverge from these PoEs, they may ultimately redefine them (§ 4.6). These models learn semantics and internal structure through specific mechanisms applied to massive datasets, replacing explicit external knowledge with data-driven inference. As such, they provide versatile methods that may exploit learnt internal structure (PoE-i) and semantic coherence (PoE-iv) to achieve both generality and precision.

4.1. *From symmetry to structure: integrating physical invariance in AI models (PoE-i)*

Classical machine learning and deep learning algorithms rely on statistical learning, i.e. they infer mappings from inputs to outputs based on patterns present in large datasets.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

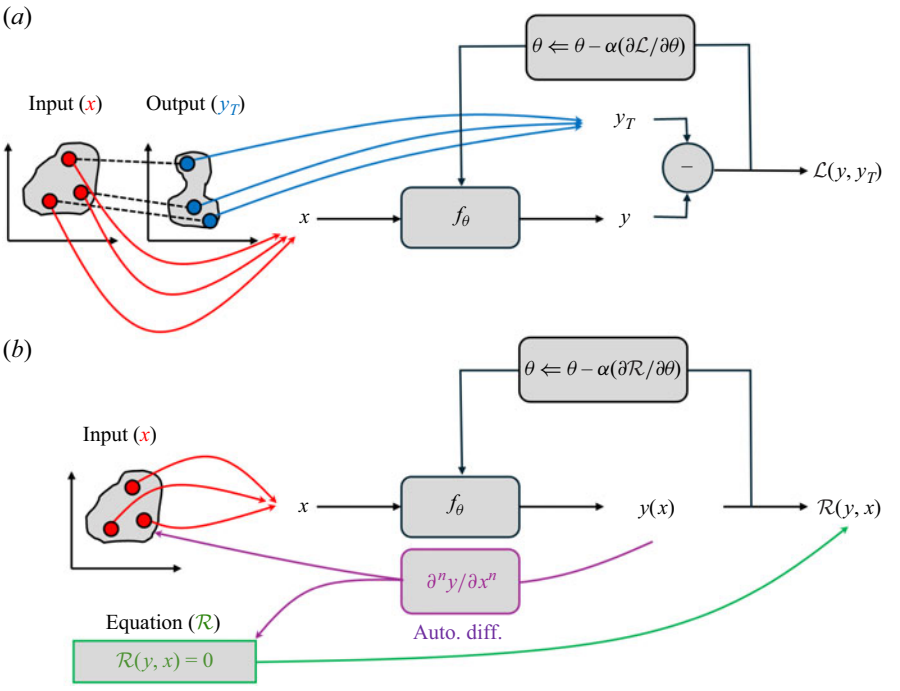


Figure 1. (*a*) Classical network, where training data (x, y_T) are sampled and fed to a parametrised function f_θ. Training is achieved by measuring a loss function $\mathcal{L}$ between the prediction (y) and the target (y_T). Network's weights θ are updated by a back-propagation algorithm and gradient descent methods. (*b*) PINN employs an equation $\mathcal{R}(y, x) = 0$. Training is achieved by computing the residual $\mathcal{R}$ of the physical equation from the predictions $y(x)$. To do so, partial derivatives $\partial^n y/\partial x^n$ are computed by automatic differentiation. Network's weights are updated using the residual $\mathcal{R}$ as physical-based loss function. Since f_θ is a continuous function of x, the output $y(x)$ is theoretically continuous in space and time (infinite resolution).

In this sense, standard models capture physical laws only implicitly and often in a weak or approximate form. This raises scepticism when applying neural networks to physics. For instance, models trained purely from data, without explicit physical structure, may violate fundamental principles such as conservation laws or symmetries (PoE-i). The lack of built-in physics is therefore widely regarded as a key limitation to both their ability to generalise beyond the training data and their widespread use in the scientific community. To address these challenges, this section will develop several approaches that may incorporate structure into AI models. First, physical equations can be explicitly enforced within AI frameworks (§ 4.1.1). Second, incorporating symmetries can help to ensure that models respect fundamental physical properties (§ 4.1.2).

4.1.1. *Physics-aware deep learning*

One of the most critical challenges in AI for CFD is ensuring that the models respect the underlying physical laws, such as conservation of mass, momentum and energy. Standard neural networks (figure 1*a*), while powerful, often lack built-in mechanisms to preserve these physical principles. This has led to the development of PINNs (figure 1*b* and table 2), which incorporate physical laws directly into the loss function. By doing so, PINNs ensure that the network's predictions are consistent with the governing equations $\mathcal{R}$ of fluid dynamics (Cai *et al.* 2021; Lu *et al.* 2021). Moreover, compared with CFD working on discretised equations, PINN approximates the solution of the continuous PDE, thus

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

providing in theory solution with an infinite resolution. Yet, in practice, this property may be limited by the sampling strategy used during training (Wu *et al.* 2023). So far, only simple flows have been tackled by PINN (Moser *et al.* 2023). To my knowledge, no examples of efficient PINN in complex 3-D flows have been shown. As a consequence, future works to improve the present framework allowing to deploy PINN in complex configurations better than classical CFD still seem illusive.

Nevertheless, several ideas behind PINN have been reused in more recent works. For instance, enforcing the physical equations into the loss function of more classical deep neural networks have shown benefits even on 3-D flows. As an example, Tompson *et al.* (2017) enforced the incompressible mass conservation equation $\boldsymbol{\nabla} \cdot \boldsymbol{u} = 0$ during training, stabilising the solution in time. It also allows a non-supervised approach, acting as a data augmentation technique, or as a manner to control the network error when hybridised with a classical solver (Illarramendi *et al.* 2020*a*). Strongly enforcing physical laws have also been proposed by Trask, Huang & Hu (2022), for instance, by exploiting exterior calculus. Similar ideas to enforce physics have been widely investigated to improve neural networks predictions of dynamical systems (Doan, Polifke & Magri 2020). For instance, Greydanus, Dzamba & Yosinski (2019) introduced Hamiltonian neural networks (HNN) and then Cranmer *et al.* (2020*a*) introduced Lagrangian neural networks (LNN) to learn and respect exact conservation laws in an unsupervised manner. In HNN, the Hamiltonian equations $\mathrm{d}q/\mathrm{d}t = \partial\mathcal{H}/\partial p$ and $\mathrm{d}p/\mathrm{d}t = -(\partial\mathcal{H}/\partial q)$ are encapsulated in the loss function when learning $\mathcal{H}$ to weakly enforce the symplectic structure of the Hamiltonian system. These techniques are built in a similar way to classical integrators for chaotic dynamical systems developed decades before, such as the symplectic integrators (Yoshida 1990). Similarly, Hernandez *et al.* (2023) proposed the concept of thermodynamics-informed neural networks to enforce a metriplectic structure of dissipative Hamiltonian systems onto the latent space network, which propagates the reduced state in an auto-regressive manner. A metriplectic structure is a framework that combines both conservative (Hamiltonian, symplectic structure) and dissipative (entropy-producing) behaviours within a single system. This structure is used to model non-Hamiltonian systems that involve energy dissipation. Note that the benefits of enforcing *a priori* physical knowledge is not limited to AI-based methods. For instance, Baddoo *et al.* (2023) revealed that standard dynamic mode decomposition can also be enriched by enforcing physical properties on the matrix A employed to learn the linear mapping that advances one snapshot of the system into the next. While usually A is searched as a low-rank matrix (Schmid 2010), Baddoo *et al.* (2023) proposed to search this matrix A as tri-diagonal, block matrices, unitary or upper triangular, depending on the physical property of the system (locality, multi-dimensional, conservative or causal).

As an example, recent research works focused on evaluating and adapting these AI methods to predict complex dynamical systems. Usandivaras *et al.* (2024*b*) employed a spatial convolutional auto-encoder (figure 2*a*) coupled with a temporal multilayer perceptron (MLP) to predict the space–time evolution of a turbulent flame in a rocket engine. The architecture resembles the work proposed by Rywik *et al.* (2023) on laminar flames. The proposed method was outperforming classical techniques (figure 2*b*), such as POD, in this turbulent case with fewer modes (20 for AI, compared with more than 100 modes for POD). However, generalisation beyond the time horizon of the training dataset was failing, because of a spectral collapse not yet elucidated. Novel architectures such as HNN and LNN might be an approach to solve this issue. Similarly, Colombo *et al.* (2023) employed a neural ODE (Chen *et al.* 2018) to constrain in time a graph neural network to predict the dynamical response of flexible drone subjected to gust (figure 2*c*). In contrast

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

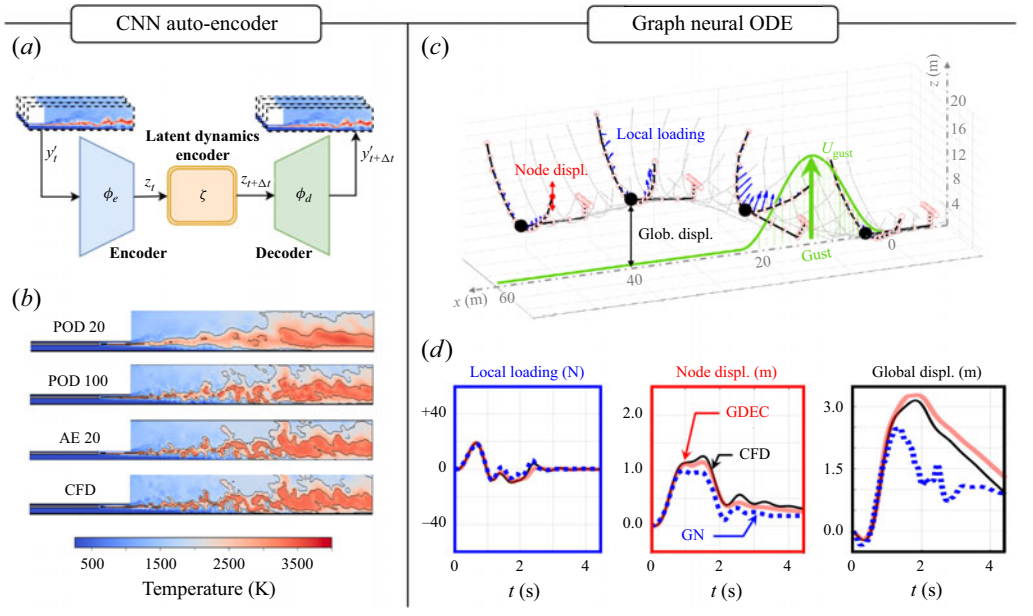


Figure 2. (*a*) CNN auto-encoder (AE) architecture to predict a dynamical system (Usandivaras *et al.* 2024*b*), where a network ϕ_e encodes the snapshot y_t, producing a latent code z_t. This code is propagated in time by a network ζ such that $z_{t+\Delta t} = \zeta(z_t)$. This new code is then decoded by the network ϕ_d, providing the high-dimensional snapshot $t_{t+\Delta t}$ of the system. (*b*) The AE algorithm shows excellent results compared with CFD, while classical POD fails to effectively reproduce the turbulent dynamics. (*c*) Sketch of the fluid–structure problem where a flexible drone flies over a gust (transverse velocity U_{gust}) (Colombo, Morlier & Bauerheim 2023). This problem is formulated through a graph, where elements of the structures are connected together. Thus, graph networks can be trained to replicate the forced fluid–structure dynamics. (*d*) Results show that a simple graph network (GN, blue) produces slight inaccuracy on node displacements, yielding large errors on the global displacements. This issue can be limited by hybridising graph networks and neural ODE (GDEC, red).

with simpler long-term-losses (LTLs) proposed by Tompson *et al.* (2017), which only learn the discretised dynamics indirectly by penalising long-term errors, neural ODE (NODE, see table 2) learns the instantaneous dynamics of the output $y(t)$:

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f_\theta(y, t). \tag{4.1}$$

NODE training is performed by differentiating directly through the time integrator, which ensures that the network learns the local dynamics rather than only trajectory statistics. As a result, long-term drift is significantly reduced, since errors remain tied to the underlying ODE structure rather than accumulating arbitrarily in time. In practice, figure 2(*d*) demonstrates that coupling NODE with graph networks (GDEC) yields more accurate predictions than classical graph networks (GNs) trained with a simpler long-term loss, illustrating the benefits of embedding physical structure within deep learning frameworks. Similar techniques to constrain prediction of dynamical systems in time have been proposed for generative networks by Xie *et al.* (2018), based on a temporal discriminator and physics-aware data augmentation.

4.1.2. *Enforcing symmetries and physical constraints in neural architectures*

Another emerging trend is designing AI architectures that inherently respect symmetries, such as rotational or translational invariance, which are crucial in fluid dynamics.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

Architecture	Domain	Symmetry / Invariance
MLP	Continuous input	None
CNN	Grid	Translation
MeshCNN	Manifold	Isometry / SO(2)
GNN	Graph	Permutation Σ_n
Transformer	Graph	Permutation Σ_n
LSTM	1D grid	Time wrapping
INR + Fourier features	Continuous coordinates	Translation and scale
NODE	Continuous time/space	Time reparametrization
HNN	Phase space (q, p)	Symplectic structure
LNN	Configuration space $(q, \dot{q})$	Variational symmetries

Table 3. Various well-known neural architectures, and their respective domain and underlying symmetry. The first part of the table is extracted from Bronstein *et al.* (2021). Note that table 1 details how aggregation functions must be adapted for each type of network presented here.

As definitions, a transformation g is a symmetry of f if f remains invariant under its action, i.e. $f(g(x)) = f(x)$. Instead, if f commutes with the action, i.e. $f(g(x)) = g(f(x))$, then f is said to be equivariant to the symmetry g. For instance, equivariant/invariant neural networks are designed to preserve the symmetry transformations of the data, ensuring that the model's outputs change in a predictable way under transformations like rotations or reflections. Note that regular convolutional layers are, by construction, translational equivariant, while pooling layers are approximately translational invariants. In fact, as mentioned by Bronstein *et al.* (2021), most network architectures are built upon specific symmetries or invariance, as reported in table 3. Note that we could extend this list with INR equipped with Fourier features (see § 4.2.3 and figure 5*b*) that are stable under translation and scale transformations, or NODE that are invariant to time reparametrisation. Similarly, previous Hamiltonian (HNN) and Lagrangian (LNN) also preserve specific structures.

Table 3 also highlights that many symmetries are not inherently encoded in standard neural architectures. For example, rotational invariance is typically absent and has often been approximated through data augmentation, where inputs are randomly rotated so that the network learns this transformation only implicitly. More recent research, however, has focused on designing architectures that explicitly enforce rotational symmetry and, more generally, group-based symmetries, leading to models with stronger inductive biases and improved generalisation. In the pioneering work of Gens & Domingos (2014), deep symmetry networks were introduced, significantly reducing the number of training examples needed compared with standard convolutional networks. These networks are designed to explicitly capture symmetries present in datasets, such as rotations in the MNIST-rot dataset (Larochelle *et al.* 2007). To illustrate the idea, consider a simple image classification task, where the input is an image X and the output is a label Y. Some transformations, such as rotations or reflections, do not change the label Y. Instead of learning a separate mapping for each transformed version of an image, deep symmetry networks exploit these transformations directly. They effectively 'share' knowledge across all transformed versions of the same image, achieving similar accuracy with far fewer training examples: for instance, only 100 samples instead of 1000 for a standard convolutional classifier on MNIST-rot (Gens & Domingos 2014).

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

In fluid mechanics, Önder & Liu (2023) enforced the symmetries governing the curved interface of multiphase flows into MLP, designed to replace volume of fluid (VOF) methods. A similar demonstration was achieved for level-set methods (Buhendwa, Bezgin & Adams 2022). Here, they simply preserve symmetry by mirroring and rotating the neural network input with regards to all symmetries, and subsequently averaging over the predictions. This is feasible only for discrete symmetries. For continuous symmetry, enforcing the symmetries of the Navier–Stokes equations by equivariant convolution operators is an emerging line of research which could help to substantially improve the generalisation accuracy of CNNs. Partly inspired by recent advances in symmetry-preserving NNs for image segmentation (Weiler & Cesa 2019), Wang *et al.* (2021) incorporated Galilean, rotation and scale symmetries into CNN models through E(2)-equivariant steerable convolutions (Weiler & Cesa 2019). Similarly, Siddani, Balachandar & Fang (2021) leveraged three-dimensional steerable CNNs (Weiler *et al.* 2018) to embed rotation equivariance into a CNN model for inferring the velocity field in the steady flow around stationary particles. Recently, Otto *et al.* (2023) proposed a seminal work dedicated to enforcing or discovering symmetries in machine learning. A complete and comprehensive work on symmetry-enforcing and gauge theory of convolutional networks has also been proposed by Weiler *et al.* (2023). Despite the promising results of symmetry-preserving CNNs for flow inference, their design and implementation add complexity, which may be limiting their fast development and adoption (Lino *et al.* 2023).

The unconstrained nature of neural networks can also lead to solutions that are physically inconsistent or lack generalisation. Beyond symmetries, other specific constraints can be integrated in the network's architecture design to address these limitations. A key approach involves Lipschitz-constrained neural networks (LCNNs, see table 1), where the Lipschitz constant of the network function is bounded. LCNNs have shown promise in applications requiring stability, monotonicity and robustness, making them particularly relevant to fluid mechanics. A function f is defined as K-Lipschitz continuous if there exists a Lipschitz constant $K \geqslant 0$ such that for all pairs of inputs x_1 and x_2, $\|f(x_2) - f(x_1)\| \leqslant K\|x_2 - x_1\|$. This equation highlights that an LCNN ensures bounded sensitivity of the neural network to perturbations in input, which is critical for systems governed by partial differential equations. It therefore limits the divergence of predicted quantities, which thus respects the bounded energy of physical systems. It also avoids the network from being too sensitive to noisy inputs $x_2 = x_1 + \epsilon$, since the output will be of the order of $\|\delta f\| = \|f(x_2) - f(x_1)\| \leqslant K\epsilon$. LCNNs have been used for robust network predictions (Béthune *et al.* 2022). 1-Lipschitz networks (i.e. the Lipschitz constant of the network is $K \leqslant 1$) trained with hinge-Kantorovitch–Rubinstein loss (Serrurier *et al.* 2021, hKR) are also well suited to learn signed distance functions (Coiffier & Béthune 2024). It is also employed for robust RL algorithms (Barbara, Wang & Machester 2024), as well as a necessary properties to ease theoretical proof associated with neural networks, in particular when dealing with generalisation and out-of-distribution capabilities (Bobbia & Picard 2024). Note however that building LCNN with spectral normalisation or Lipschitz regularisation may increase training time, as well as reducing the network expressivity for highly nonlinear or sharp behaviours. To my knowledge, LCNN has not yet been used in fluid mechanics. I however believe that LCNN will attract some attention to ensure robustness and generalisation capabilities for deep networks in physics in the near future, for both regression and control tasks.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

4.2. *Scale separation in practice: from multi-scale networks to multi-fidelity learning (PoE-ii)*

4.2.1. *Learning across scales*

Fluid flows are characterised by their inherent nonlinearity and multi-scale behaviour, particularly in high-Reynolds-number regimes where turbulence dominates. In the Navier–Stokes equations, the nonlinear term $(\boldsymbol{v}\cdot\boldsymbol{\nabla})\boldsymbol{v}$ governs interactions between these scales. Traditional numerical methods discretise the gradient term using numerical stencils, regardless of the specific physical phenomena or scales present. To address this, the underlying computational mesh must be carefully designed in advance to capture flow scales. Estimation of these scales *a priori* is thus required, mostly based on the user's experience on the considered problem. Adaptive mesh refinement (AMR) was developed (Berger & Oliger 1984) as an iterative process to dynamically refine the mesh in regions where smaller scales emerge, reducing the human intervention required for the mesh improvement. However, when the required resolution is computationally prohibitive, approximations are necessary. These include turbulence models to represent unresolved scales, or artificially thickened representations of shock waves (Lodato 2019) or thin flames (Colin *et al.* 2000), which are activated only when a manually designed sensor indicates they are needed. Traditionally, such sensors rely on classical operators (e.g. gradient, divergence, curl) to detect relevant physics and generate metrics for model adjustment. However, these approaches often require complex, *ad hoc* formulations to ensure well-posedness, generalisation and efficiency.

Here, AI offers a compelling alternative by leveraging flow topology directly. Convolutional neural networks (CNNs), for instance, excel at detecting and classifying flow structures, even on unresolved or coarse grids. Deep networks have demonstrated significant improvements in robustness and accuracy for tasks like shock detection without user intervention (Beck *et al.* 2020). Similarly, they have been employed to model unresolved structures, such as those encountered in thin turbulent flames (Xing & Lapeyre 2023). In essence, AI, drawing from advancements in computer vision, enables flow simulations to adapt to the observed flow topology rather than rigidly applying numerical stencils agnostic to the underlying phenomena. This capability is particularly advantageous for irregular and noisy flow patterns, which often appear on coarse grids. Building on this principle, Alguacil *et al.* (2022) successfully trained AI models on LBM data to capture the behaviour of acoustic waves propagating through complex media. Their models exhibited good performance, especially on coarser grids where traditional numerical schemes tend to fail. Similarly, Bar-Sinai *et al.* (2019) introduced data-driven discretization to learn optimised approximations of PDEs working on coarse grids. Drozda *et al.* (2023) used deep learning to learn an optimised Taylor–Galerkin convection scheme by examining both the local physical field and the local mesh topology.

4.2.2. *Learning multi-scale dependencies in fluid flows*

Despite their promise, early studies also highlight the difficulty of neural networks, such as CNN, in addressing multi-scale problems. First, because CFD simulations typically use unstructured meshes while CNNs require data on a uniform Cartesian grid, a preprocessing interpolation step is needed to convert the CFD fields into a format suitable for the CNN. This interpolation often leads to large errors in the flow containing small-scale variations, as encountered for instance at the trailing edge of aerofoils. In this regard, Catalani *et al.* (2023) evaluate this limitation of convolutional layers for multi-scale problems, such as transonic aerodynamics where smooth pressure variations coexist with sharp shock waves. Here, graph neural networks (GNNs) show better results, mainly

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

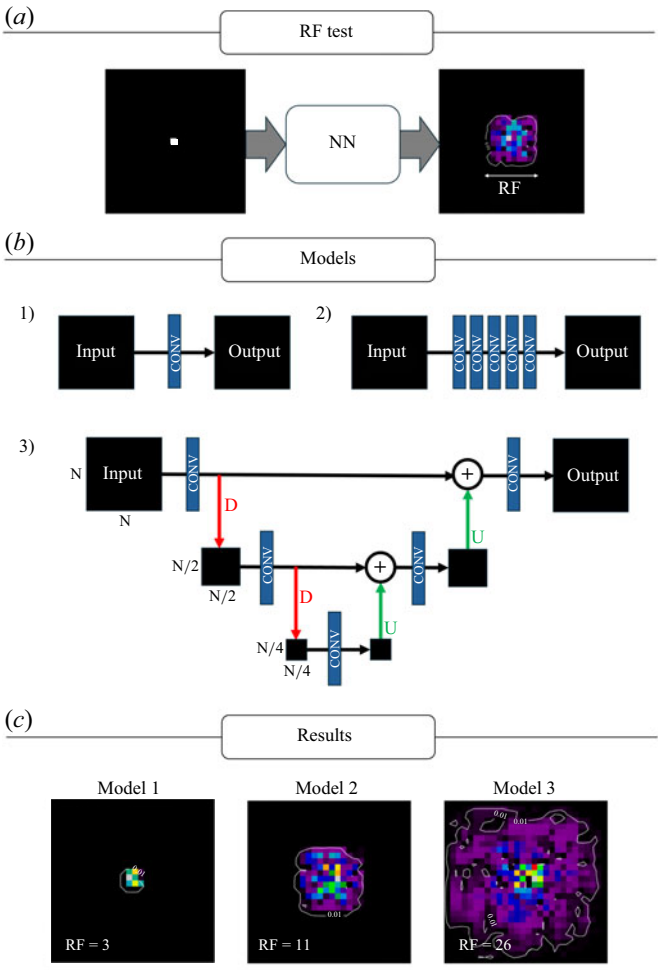


Figure 3. (*a*) Sketch of the experiment to evaluate empirically the receptive field (RF) of a neural network (NN). The network is randomly initialised and the central pixel of a constant image is perturbed, here in white. The size of the non-uniform zone of the output highlights the RF value. (*b*) This test is done on three models: (1) a network with one convolutional layer with kernel size 3×3; (2) a network with five sequential convolutional layers with kernel size 3×3; and (3) a UNet architecture containing also five convolutional layers with kernel size 3×3, but dispatch over three branches, acting at different resolutions (N, N/2 and N/4). Convolutional layers are marked with a blue rectangle, D refers to down-sampling, U to up-sampling and $+$ to concatenation. Output of the three models are given in panel (*c*), showing clearly the advantage of UNet to provide larger RF. This empirical RF value is given for the three models: 3, 11 and 26.

because of their capability to perform convolutions directly on the CFD mesh, thus eliminating interpolation errors.

Nevertheless, CNNs operate on the principle of locality, characterised by their receptive field (RF, figure 3). The RF defines the extent of neighbouring pixels used for prediction, effectively limiting the network's capacity to handle large-scale interactions that exceed the RF length (Cheng *et al.* 2021). Figure 3(*a*) presents a test to obtain empirically the RF length of a network. It consists in perturbing the central pixel (here in white) over a constant input (here in black). Since the RF only depends on the network's architecture and not on the actual values of its weights, the network can be randomly initialised. The output will show a region of non-uniform pixels, corresponding to pixels actually influenced by

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

the perturbed white input: the RF length is the size of this region, a kind of 'angle of view' of the network. It serves as typical length scale controlling the capability of the network architecture to search for interactions: two pixels in the input image separated by a length $\lambda > RF$ will be treated separately by the network, while if $\lambda < RF$, the network will see them both simultaneously, allowing to capture possible interactions to predict the output. This notion is therefore crucial for AI in physics, but not sufficiently considered or investigated in practical applications.

A general rule is that for feedforward CNN using C convolutions with kernel size $K_c \times K_c$, the RF is theoretically

$$RF = 1 + \sum_{c=1}^{C} (K_c - 1). \tag{4.2}$$

For example, a simple CNN with one convolutional layer (model 1) used $C = 1$ and $K = 3$, which yields $RF = 1 + 1 \times 2 = 3$, while increasing its depth to 5 convolutional layers (model 2) gives only $RF = 1 + 5 \times 2 = 11$, theoretical values well captured by the empirical test provided in figure 3(*a*). As a consequence, achieving a large receptive field with such a simplistic structure requires very deep networks (large C), which increases the risk of overfitting, leads to a high memory footprint and makes the model more prone to vanishing gradient issues. Consequently, while large RF is beneficial in physics to capture long-distance relationships between inputs, figure 3 and equation (4.2) demonstrate that constructing such networks using plain feedforward architectures often necessitates either extremely deep networks or large kernel sizes, which can lead to inefficiencies.

However, for CNNs, an effective alternative is to employ multi-scale architectures that use downsampled branches, obtained by the pooling operations that transform an input at resolution N into a smaller resolution $N/2$, thus discarding the smallest wavelength of the image and keeping only the largest structures. This typical architecture extends the RF without requiring excessively deep networks. A notable example is the well-known UNet architecture, depicted in figure 3(*b*) (model 3), which incorporates multiple branches (here 3) dedicated to progressively downsampled versions of the input. This design allows UNet to effectively handle multi-scale phenomena, as illustrated in figure 4, where a comparison is provided between a feedforward CNN (MONO), a UNet CNN (MULTI) and the reference CFD solution. Note that the two networks have the same number of parameters and only differ by their architecture. This highlight the effectiveness of employing multi-scale networks in CFD. In addition, recent works (Fanaskov 2024; Gode *et al.* 2024) have bridged the gap between multi-scale neural architectures, like UNet, and traditional multigrid methods, commonly used in linear solvers for instance.

In contrast, graph neural networks face unique challenges when attempting to extend their RF or multi-scale capabilities. Simply scaling the depth or width of GNNs is insufficient due to optimisation difficulties, such as representation oversmoothing (node features converge towards similar values) or oversquashing (long path to propagate information compresses relevant signals and reduces their influence) (Li, Han & Wu 2018; Rusch, Bronstein & Mishra 2023). Unlike CNNs, pooling-based approaches are not universally applicable in GNNs, making it difficult to adapt the effective UNet-like architectures (as model 3 in figure 3) for graph-based domains. This limitation suggests the need for more sophisticated strategies to address the multi-scale nature of fluid flows on general grids (Wu *et al.* 2021). A promising solution to this problem is demonstrated in the seminal GraphCast framework (Lam *et al.* 2023), which employs a multi-mesh message-passing strategy to overcome the limitations of traditional GNNs.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

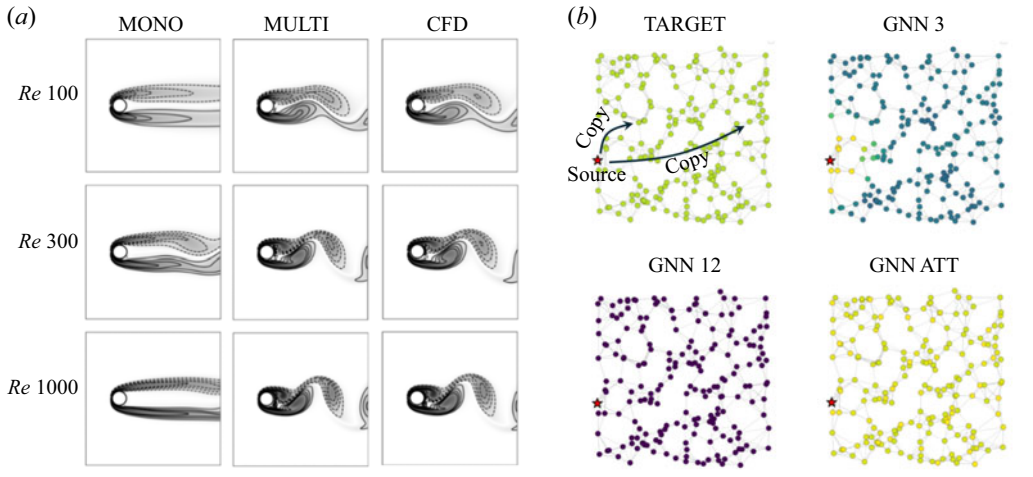


Figure 4. (*a*) Results obtain by a deep neural network to solve the Poisson equation embedded in an incompressible solver (Illarramendi *et al.* 2020*b*). Training is performed on non-viscous flows. Test results on viscous flows at various Reynolds numbers ($Re = 100$, 300 and 1000) are presented for a feedforward CNN (MONO), UNet architecture (MULTI) and the reference CFD solutions. Note that MONO and MULTI have the same number of weights, they only differ by their architecture, revealing the excellent performance of multi-scale networks like UNet in CFD. (*b*) Experiment on a test consisting in copying a value of a source node ($\star$) in all other nodes of an unstructured mesh. Results are provided for a GNN with 3 layers (GNN 3) and 12 layers (GNN 12), as well as with a graph network with global attention (GNN ATT). This illustrates the locality of shallow network (GNN 3), the difficulty of training large graph networks (GNN 12) and the excellent behaviours of graph when equipped to capture long-distance relationships (GNN ATT).

By leveraging multiple resolution scales within the graph structure, GraphCast achieves superior performance in capturing multi-scale behaviours while maintaining stability and efficiency. Similarly, the classical message-passing (MP) strategy employed in graph networks can be replaced by attention mechanisms (Hussain, Zaki & Subramanian 2022) that help capturing long-distance relationships between nodes (see attention mechanism in table 1). While MP approaches mimic numerical schemes on unstructured meshes, they strongly suffer from locality and oversmoothing. This is illustrated in figure 4(*b*) on a simple example consisting in copying the value of a source node ($\star$) to all nodes of the unstructured mesh. While a shallow graph network (GNN 3, with 3 MP layers) exhibits a limited zone of action, highlighting its difficulty of capture long-distance relationships, deeper graph networks (GNN 12, with 12 MP layers) completely fail. This can be overcome by a graph network with global attention mechanism (GNN ATT) which learn the correlation between each nodes of the mesh, irrespective of their distance. Results in figure 4(*b*) shows that GNN ATT successively address the task, while other MP-based graph networks fail. Note that since message-passing is achieved on all edges of nodes, its cost scales linearly, as $\mathcal{O}(kN)$, where N is the number nodes and k the number of edge per node. However, global attention requires the correlation between all nodes, thus scaling quadratically, as $\mathcal{O}(N^2)$, which can explode for large meshes. This difficulty can be mitigated since attention mechanisms rely on dense matrix multiplications that are well-optimised and highly parallelised on GPUs. These developments highlight the importance of specialised architectural innovations to tackle the inherent multi-scale challenges in fluid dynamics, particularly when working with diverse grid structures, irregular domains and long-distance relationships between variables.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

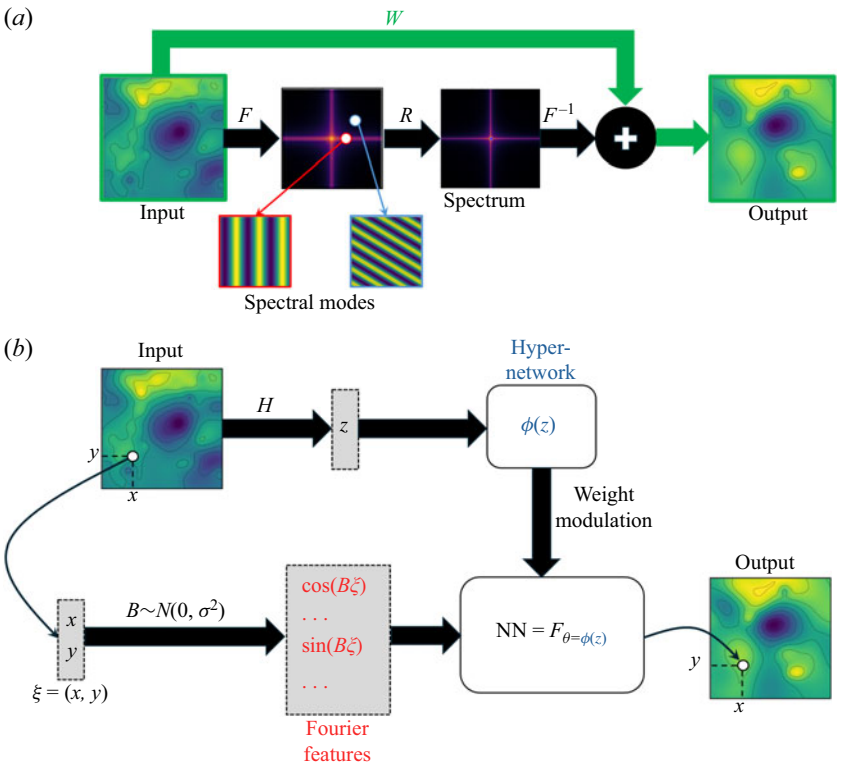


Figure 5. (*a*) Illustration of a FNO layer, which combines a learned linear operator mixing channels (black arrows, matrix R) in the Fourier space with a local linear map in physical space (green arrows, matrix W) acting through a skip connection. This allows to incorporate long-distance relationships between pixels, since spectral modes are specifically designed to separate and transform features at each scale. (*b*) INR implicitly learns the long-distance relationship by outputing directly the desired quantity $F_\theta(x, y)$ at a query point (x, y). However, training can be limited by spectral bias, which makes high frequencies more difficult to learn. To avoid this, spectral embeddings are incorporated (Sitzmann *et al.* 2020). A drawback of INR is its difficulty to input field, which requires specific compression and representation techniques (latent variable z), hyper-networks (ϕ) that modulate the weight $\theta = \phi(z)$ of the main network, and meta-learning strategies.

4.2.3. *Overcoming scale separation with neural operators*

While previous graph architectures explicitly models interactions between nodes, emerging architectures like DeepONet (Lu, Jin & Karniadakis 2019), Fourier neural operators (FNO, figure 5*a*) (Li *et al.* 2021), wavelet neural operators (WNO) and more recently implicit neural representations (INR, figure 5*b*) with Fourier features (Sitzmann *et al.* 2020; Luigi *et al.* 2023; Yin *et al.* 2023) are advancing the ability of AI models to tackle multi-scale challenges in fluid dynamics. Operator learning frameworks (see table 2) are particularly well suited for tasks involving differential operators and physics-based systems. These architectures map functions to functions, making them ideal for modelling complex phenomena where traditional methods struggle.

On one hand, FNO and related spectral approaches operate directly in the frequency domain, enabling more global and efficient representations of fluid dynamics. By leveraging Fourier transforms, FNOs extend traditional CNNs to learn operators in spectral space, effectively capturing the non-local interactions characteristic of multi-scale fluid phenomena (figure 5*a*). However, FNOs face a trade-off: the number of retained Fourier modes directly impacts both computational cost and accuracy, as each mode corresponds to a specific spatial scale. While standard FNOs include a local linear map in physical

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

Test case	GNO	DeepONet	FNO	WNO
Burger	6.15 %	2.15 %	**1.60 %**	1.75 %
Darcy (rect.)	3.46 %	2.98 %	1.08 %	**0.84 %**
Darcy (tri.)	—	2.64 %	—	**0.77 %**
Navier–Stokes	—	1.78 %	1.28 %	**0.31 %**
Wave-advection	—	**0.32 %**	47.7 %	0.62 %

Table 4. Mean L^2 relative error between the true and predicted results using various operator learning approaches, among which Graph Neural Operator (GNO), DeepONet, Fourier Neural Operator (FNO) and Wavelet Neural Operator (WNO). Best results are highlighted in bold. Table is extracted from Tripura & Chakraborty (2023).

space (green arrows in figure 5*a*), they remain limited in their ability to represent strongly localised or spatially varying transformations. Wavelet neural operators (WNOs) address this limitation by using wavelet transforms to provide simultaneous localisation in space and scale, enabling more flexible treatment of both global and local dynamics (Tripura & Chakraborty 2023). Recent advances have further refined these concepts. For example, Lütjens *et al.* (2022) introduced multi-scale neural operators that separate scales through a filtering process similar to LES in CFD. These models achieve quasi-linear complexity, i.e. $\mathcal{O}(N\log N)$.

Building on these ideas, Wang *et al.* (2024) combined FNOs with UNet-like architectures, achieving improved accuracy across scales in turbulent flow simulations. This hybrid approach outperformed both classical subgrid models, such as dynamic Smagorinsky (Germano *et al.* 1991; Meneveau & Lund 1997) and WALE (Nicoud & Ducros 1999), as well as standard AI models based on plain FNOs. Table 4 compares the performance of various operator learning techniques, such as graph neural operators (GNO), DeepONet, FNO and WNO, across diverse test cases. These include Burger's equation, Darcy's equation on rectangular (rect.) and triangular (tri.) domains, 2-D Navier–Stokes equations, and a 1-D time-dependent wave advection problem. Note that standard FNO (Li *et al.* 2021) cannot handle complex domains like the triangular test case for Darcy's equation. Results highlight the benefits of addressing spatial scale challenges when solving flow-related PDEs, with spectral approaches like FNO and WNO consistently outperforming others.

However, implicit neural representations (INRs) provide a framework for representing individual continuous functions using neural networks (figure 5*b*). In this approach, an MLP maps spatial coordinates $\xi = (x, y)$ to field values $F_\theta(\xi)$, yielding a continuous resolution-independent representation of the solution. To handle families of flows depending on physical or geometric parameters (e.g. boundary conditions, operating points, geometry descriptors, denoted z in figure 5*b*), the network weights θ are generated by a hypernetwork ϕ conditioned on these parameters z. Concretely, the hypernetwork takes z as input and outputs a specific set of weights $\theta(z)$, thereby defining a dedicated spatial function $F_{\theta(z)}(\xi)$. Training is performed by backpropagating the loss $\mathcal{L}$ through the INR into the hypernetwork, i.e. $\partial\mathcal{L}/\partial\phi = (\partial\mathcal{L}\partial\theta)(\partial\theta/\partial\phi)$, and only the hypernetwork parameters ϕ are updated to learn how solution functions vary across the parameter space. This design differs from simply concatenating parameters with spatial coordinates: instead of forcing a single network to represent all solutions at once, the hypernetwork learns a structured family of solution functions, improving expressivity and generalisation across regimes. As such, this formulation enables parametrised PDE learning and high-fidelity reconstruction of flow fields.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

Nevertheless, vanilla INRs suffer from spectral bias, which makes high-frequency components more difficult to learn. By leveraging Fourier features (spectral embeddings), INRs can circumvent this issue and effectively capture multi-scale structures (Grattarola & Vandergheynst 2022; Catalani *et al.* 2024), making them a highly efficient alternative to traditional grid-based methods. To do so, spectral embeddings do not feed $\xi = (x, y)$ directly into the network, but instead, the input is lifted into a higher-dimensional Fourier embedding $\gamma(x) \in \mathbb{R}^{2N}$, defined as

$$\gamma(\xi) = \left[\cos(2\pi b_1^\top \xi), \ldots, \cos(2\pi b_N^\top \xi), \sin(2\pi b_1^\top \xi), \ldots, \sin(2\pi b_N^\top \xi)\right], \tag{4.3}$$

where each frequency vector $b_k \in \mathbb{R}^2$ is sampled independently from a Gaussian distribution $b_k \sim \mathcal{N}(0, \sigma^2 I_2)$. Consequently, each new component acts like a Fourier mode at some random frequency and, together, they form a random Fourier basis. Feeding these oscillatory inputs to the MLP gives the network the capacity to represent high-frequency structure much more effectively. To go further, Tancik *et al.* (2020) employs the neural tangent kernel (NTK) theory, which describes how a large neural network f_θ behaves like a kernel method, with kernel $K(\xi, \xi') = \nabla_\theta f_\theta(\xi)^T \nabla_\theta f_\theta(\xi')$ for two inputs ξ and ξ'. They show that with Fourier features $\gamma(\xi)$, the NTK behaves like a shift-invariant kernel $K(\xi, \xi') \approx K(\xi - \xi')$. As a result, the learning dynamics becomes translation-equivariant: the network learns spatial patterns independently of their absolute location, without this symmetry being explicitly imposed. Therefore, this bridges the gap between our PoE-i on invariance and PoE-ii on scale separation. It yields that training such a network is equivalent to doing kernel regression with a well-behaved kernel, giving predictability, interpretability and generalisation guarantees.

The remaining hyper-parameter is the variance σ^2, which controls the characteristic spectral scale emphasised during training. Although frequencies are sampled from a broad Gaussian distribution, NTK analysis shows that the resulting kernel concentrates learning around spatial frequencies $k \sim \sigma$, while attenuating both lower and higher frequencies. As a result, a single σ value biases learning towards a narrow frequency band. Multi-scale Fourier encodings, obtained by combining multiple values of σ (e.g. $\sigma = (1, 5)$ in figure 6*b*), effectively superimpose kernels acting at different spectral scales, enabling faster and more balanced learning of both global and localised features. Figure 6(*b*) illustrates the advantage of multi-scale spectral encoding on a simple 1-D signal, where the choice $\sigma = (1, 5)$ significantly outperform single-scale encodings ($\sigma = (1)$ or $\sigma = (5)$), demonstrating superior ability to capture complex features.

Finally, INRs equipped with Fourier features exhibit near-resolution invariance and enable rapid training on coarser grids without sacrificing accuracy. This property has been successfully applied to realistic aerodynamic scenarios by Catalani *et al.* (2024), who demonstrated that multi-scale INR methods outperform classical graph neural networks (MGN) and Gaussian processes in terms of accuracy, robustness and resolution invariance on 2-D transonic aerofoils and 3-D wings (see figures 6*c* and 6*d*). Recent advancements in INR techniques further enhance their capabilities. For instance, local neural fields equipped with attention mechanisms, such as the AROMA architecture (Serrano *et al.* 2024), improve multi-scale predictions by focusing computational resources on critical regions. Additionally, INRs have been adapted to operate on non-Euclidean manifolds with intrinsic neural fields (Koestler *et al.* 2022), expanding their applicability to complex geometries in fluid dynamics and beyond.

These methods are relatively recent and are not yet fully mature when applied to complex, multiscale three-dimensional fluid flows. A deeper theoretical understanding of how these innovative deep learning networks capture the nonlinear interactions of multiple scales is essential. For instance, in INRs, Fourier features are currently set

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

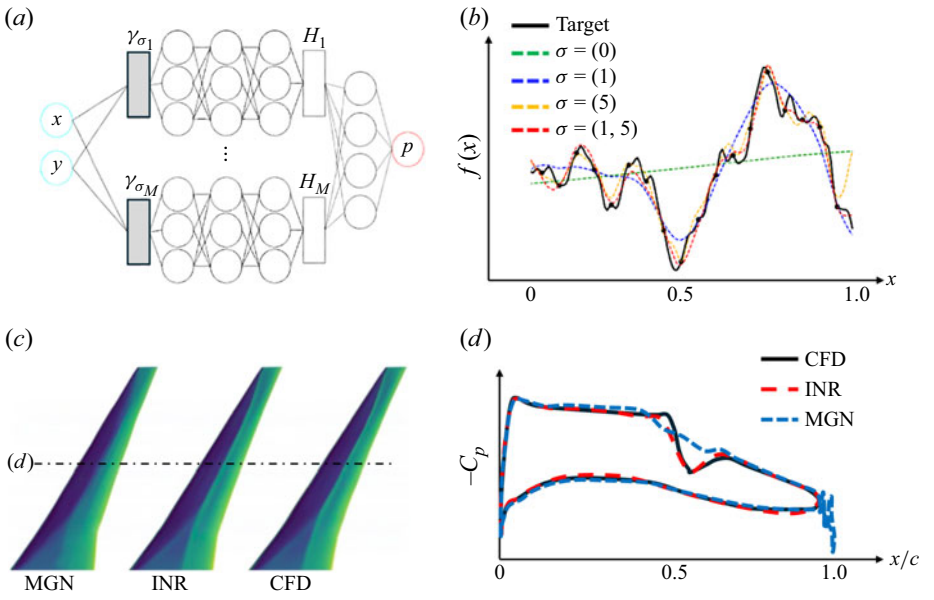


Figure 6. (*a*) Illustration of a multi-scale INR architecture with various embedding vectors γ_σ. (*b*) Experimental results for fitting a single signal using a 3-layer INR with different Fourier encodings. This experiment can be reproduced from our Git repository. (*c*) Results obtained by CFD, graph network (MGN) and INR on 3-D wing pressure predictions, with (*d*) a quantitative comparison on the pressure coefficient C_p predictions.

manually without clear guidelines grounded in the physical scales of the flow. Developing such guidelines could significantly enhance their applicability and performance. Recent alternatives, such as the instant-NGP approach proposed by Müller *et al.* (2022), encode coordinates into several grids of different resolutions. This method offers low-storage solutions for multi-resolution highly detailed volumetric data, suggesting potential pathways for improvements. Adopting a more structured framework with explainable guidelines, similar to the well-established practices in CFD, could lead to faster training, improved robustness and broader adoption within the fluid dynamics community.

4.2.4. *Capitalising on multi-fidelity data*

In computational fluid dynamics, achieving high-resolution simulations is often computationally expensive. Multi-fidelity methods aim to bridge the gap between low-fidelity (LF), computationally inexpensive models, and high-fidelity (HF), accurate simulations. These methods leverage the inherent distinctions between models, often related to the scales they resolve. For example, potential flow models assume zero vorticity, effectively eliminating viscosity and thus avoiding the formation of thin boundary layers. This simplification significantly reduces computational cost by bypassing the need to resolve small-scale flow features present in realistic scenarios, in contrast with high-fidelity simulations like RANS or LES. In other words, exploiting multi-fidelity strategies is by essence exploiting the scale separation principle (PoE-ii).

The core idea behind multi-fidelity methods is to exploit the strengths of various models: leveraging LF tools for rapid exploration of the design space, while reserving HF simulations for refining regions of interest. These approaches are particularly valuable in uncertainty quantification (Peherstorfer, Willcox & Gunzburger 2018) and optimisation problems (Thelen *et al.* 2022) in high-dimensional spaces, where the cost of exclusively

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

using HF simulations would be prohibitive. Unlike multi-level methods, which rely on a strict hierarchy between models, multi-fidelity approaches are more flexible, accommodating heterogeneous models without assuming a predefined relationship. Several strategies have emerged for constructing multi-fidelity frameworks, including: (i) incorporating information from HF simulations to refine or correct LF predictions; (ii) fusing outputs by combining predictions from LF and HF models to generate an integrated solution; and (iii) selective refinement, by using HF models only in critical regions or where higher accuracy is essential. The main goal of these strategies is to ensure that the final output achieves the same level of accuracy as if only HF models were used, while significantly reducing computational cost.

AI in CFD faces challenges similar to those encountered in uncertainty quantification (UQ) and optimisation tasks, particularly the need to collect expensive data samples in high-dimensional spaces. Consequently, integrating, adapting or reimagining multi-fidelity strategies within AI becomes essential. The potential benefits are twofold: (i) efficient exploration and exploitation of the high-dimensional input space by leveraging models with varying fidelity levels; and (ii) effective handling of flow scales, with different models naturally capturing specific flow features depending on their fidelity. A compelling example of this concept is seen in multi-scale INRs previously described in § 4.2.3 and figure 6(*a*), which fall into the fusion category by combining outputs from multiple scales σ_k: as illustrated in figure 6(*a*), a shared MLP is applied to inputs encoded with distinct Fourier features γ_{σ_k}, producing scale-specific representations H_k, which are then fused through a final layer to predict the target quantity P.

In deep learning, multi-fidelity strategies are extensively explored in the fields of domain adaptation (Farahani *et al.* 2020; Liu *et al.* 2022*b*, DA) and transfer learning (Bozinovski & Fulgosi 1976; Pan & Yang 2010, TL). These approaches enable the transfer of knowledge across different fidelity levels and scales, allowing for rapid and accurate predictions while minimising the need for costly high-resolution simulations for every new problem. For instance, Liu & Wang (2019) introduced a physics-constrained neural network (PCNN) trained on LF data and subsequently refined using a secondary network fed with HF data to learn the discrepancies between the two fidelity levels. This framework was successfully demonstrated in a two-dimensional heat-transfer and phase transition problem. Similarly, Minisci & Vasile (2013) employed artificial neural networks to correct aerodynamic forces predicted by an LF model using HF CFD data as a reference.

Among these strategies, transfer learning is particularly promising. It involves training a model on abundant LF data and then fine-tuning it on limited HF simulations or experimental datasets, thereby enhancing generalisability and reducing computational costs. This approach has been applied in various physics-based scenarios. Usandivaras *et al.* (2024*a*) demonstrated transfer learning in turbulent combustion simulations for rocket engines (figure 7*a*). By training on hundreds of under-resolved inaccurate LES and fine-tuning on a small set of highly resolved LES, the method achieved strong generalisation to unseen configurations (figure 7*c*). Note that figure 7(*a*) clearly illustrates the inherent scale differences between LF and HF simulations. Similarly, Lyu *et al.* (2023) employed transfer learning with FNO to perform multi-fidelity predictions on heat transfer and 3-D aerodynamic applications.

4.3. *Towards sparse AI in fluid dynamics (PoE-iii)*

Sparsity is a central principle underlying the effectiveness of neural networks in capturing physical phenomena, particularly in fluid mechanics. In contrast to generic machine

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

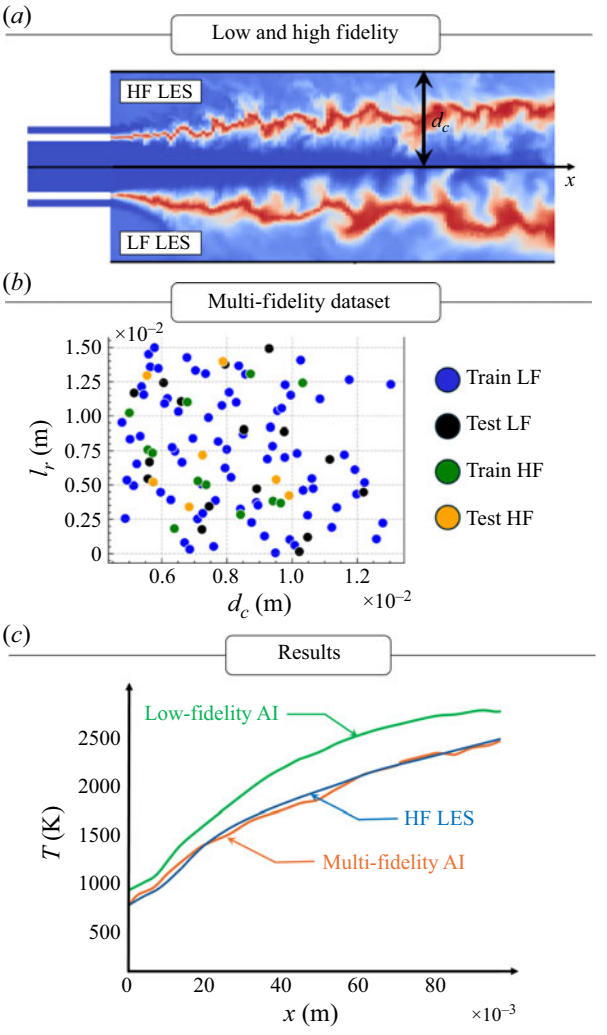


Figure 7. (*a*) Multi-fidelity configuration where fine (HF, top) and coarse (LF, bottom) LES have been performed by Usandivaras *et al.* (2024*a*). (*b*) A 2-D portion of the dataset, highlighting the LF and HF datasets when the recess length (l_r) and chamber diameter (d_c) are varied. (*c*) Results reveal that the LF AI model (green) is inaccurate compared with the true HF LES data (CFD, blue), while the multi-fidelity model trained on LF and fine-tuned on HF is in good agreement with HF LES data.

learning tasks, physical systems often exhibit intrinsic structure: only a subset of modes, interactions or spatial features dominate the dynamics. Exploiting this sparsity not only reduces the number of required parameters but also improves interpretability and generalisation, making neural networks more effective in physics-informed tasks.

Nevertheless, while sparsity appears as a natural feature of classical mathematical formulations, deep learning methods are, however, often over-parametrised. This offers flexibility to tackle complex tasks but comes at the cost of intensive training requirements, potential generalisation issues and lack of interpretability. To address this, the AI community has widely adopted sparsity-promoting techniques, such as weight-based regularisation, dropout (Gal & Ghahramani 2016) and more recently, constraints based on Lipschitz continuity (Liu *et al.* 2022*a*) or physics-informed priors (Nabian & Meidani

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

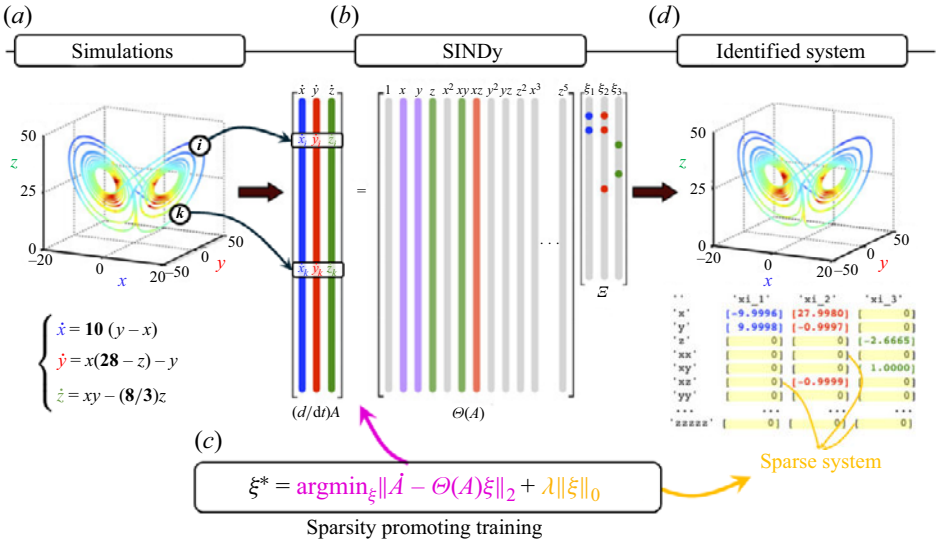


Figure 8. Algorithm SINDy proposed by Brunton *et al.* (2016*a*) for sparse identification of dynamical system. Here, illustration on the 63 Lorentz system (*a*) with $\sigma = 10$, $\rho = 28$ and $\beta = 8/3$. SINDy search for a linear model Ξ, where features are given by a function library $\Theta(A)$. (*c*) Training is achieved by minimising the error $\|\dot{A} - \Theta(A)\xi\|_2$, in addition to a regularisation term $\lambda\|\xi\|_0$ to enforce sparsity. It yields identified systems (panel *d*), which are both accurate and sparse, fostering interpretability. Here, true coefficients are well approximated: $\sigma^* \approx 9.9998$, $\rho^* \approx 27.998$ and $\beta^* \approx 2.6665$. Figure is modified from Brunton *et al.* (2016*a*).

2020). As a consequence, the primary goal of these techniques is not necessarily to improve explainability, but rather to mitigate numerical difficulties and enhance model performance, particularly for out-of-distribution inputs.

One of the most prominent examples is sparse identification of nonlinear dynamics (SINDy, see table 2) introduced by Brunton *et al.* (2016*a*). In this framework, illustrated in figure 8, governing equations of fluid flows are discovered by enforcing sparsity on a library of candidate functions Θ. Example is shown here on the well-known Lorentz 63 system (figure 8*a*), where a linear model is searched as $\dot{A} = \Theta(A)\Xi$ (figure 8*b*). By penalising non-essential terms (figure 8*c*), SINDy extracts a minimal set of active terms sufficient to describe the dynamics (figure 8*d*), providing physically interpretable models with a fraction of the data typically required by dense neural networks. Interestingly, Zhang & Schaeffer (2019) have demonstrated the convergence of such an algorithm towards a local minimiser of an unconstrained ℓ_0-penalised least-squares objective, i.e. where ℓ_0 counts the number of non-zero components. It exemplifies that sparsity is essential, not only for interpretability but also for credibility and reliability. This has also been formalised in a unified framework for sparse regression named SR3 (Zheng *et al.* 2018). Extensions of these approaches to PDEs, such as PDE-FIND, highlight that sparsity is not merely a computational convenience, but a pathway to uncovering the underlying physics from real-worlds measurements (Rudy *et al.* 2017).

Beyond explicit equation discovery, sparsity principles emerge in network architectures themselves. For instance, the U-Net architecture (Ronneberger, Fischer & Brox 2015), widely used in fluid simulations for super-resolution and subgrid modelling (Bar-Sinai *et al.* 2019; Xing & Lapeyre 2023), can be interpreted as a form of learned multi-resolution analysis, similar to the wavelet transforms (Mallat 2009). In wavelet theory, physical signals are represented with only a few significant coefficients, capturing coherent structures across scales, which make them effective for identification and denoising

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

tasks. The convolutional and skip connections in U-Net naturally focus on essential features, while discarding redundancies, imposing sparsity in the representation of flow fields. This perspective explains the remarkable ability of such architectures to generalise across different Reynolds numbers and flow regimes despite being trained on limited datasets.

Sparsity also plays a role in graph-based and spectral neural networks for fluid dynamics. Graph neural networks (GNNs) often leverage sparse connectivity mirroring the underlying physics (e.g. neighbouring interactions in a mesh), which reduces computational cost, while respecting locality. Spectral neural operators, such as FNO (Li *et al.* 2021), can be interpreted as learning sparse representations in Fourier space, where only a few modes dominate long-range interactions. Similarly, wavelet-based neural operators (Lei & Li 2025) exploit sparsity in multi-scale decompositions, allowing efficient learning of complex spatiotemporal phenomena in turbulence and multi-physics flows.

Additionally, sparsity can emerge in latent representations, widely employed in deep learning pipelines for flow physics. Auto-encoders (an example from Usandivaras *et al.* 2024*b* is shown in figure 2*a*) or latent-space neural operators often compress high-dimensional flow data into low-dimensional embeddings capturing only the dominant degrees of freedom (Lusch, Kutz & Brunton 2018). Such representations retain essential structures like vortices and coherent patterns, while discarding noise and less relevant modes, echoing the classical notion of model reduction in fluid mechanics. As an example, figure 2(*b*) reveals that the turbulent flame dynamics can be approximated by only 20 nonlinear modes when using sparse latent representations, while classical linear POD methods require more than 100 modes to be accurate. Yet, the explicability and the underlying structure of this nonlinear latent space is still unclear today, which suggests that sparsity alone may be not sufficient for interpretability.

Finally, the concept of sparsity is deeply connected to entropy in information theory: sparse vectors with only a few non-zero components exhibit low entropy, whereas random variables requiring numerous non-zero coefficients have high entropy. This relationship underlies various techniques that minimise entropy to encourage models to retain only the most essential information (Tartaglione *et al.* 2021; Barbiero *et al.* 2022), often with the goal of enhancing explainability. For instance, the well-known L1 regularisation can be interpreted as reducing entropy under a Laplace prior distribution on the network weights, which actually forces most weights to be zeros, i.e. promoting sparse networks.

In summary, sparsity is a multifaceted principle in neural networks for fluid mechanics. It manifests through: (i) explicit regularisation, as in SINDy; (ii) architectural design, as in U-Nets and multi-resolution networks; or (iii) effective representation of high-dimensional data, as in latent-space compression. Across all these perspectives, sparsity enhances effectiveness, interpretability and generalisation, reinforcing its status as a fundamental pillar for effective AI in CFD.

4.4. *Meaning beyond data: embedding semantics into AI systems (PoE-iv)*

As discussed in § 3.2, a fundamental aspect of effectiveness in mathematics and physics is their semantic structure (PoE-iv), which enables explanation that goes beyond mere symbols (syntax) to encompass the relationships and underlying meanings (semantics). Even though sparsity may foster explicability, this semantic depth is frequently noted as lacking in artificial intelligence, where deep neural networks, for instance, process inputs and features without any inherent understanding of their physical meaning. Consequently,

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

treating deep networks as 'black boxes' often leads to uncertainty and scepticism, especially in scientific applications.

An interesting idea for AI is to hybridise the current connectionist approaches (i.e. deep networks) with AI methods that inherently encapsulate semantics to get the best of both worlds. This latter approach is known as symbolic AI, which refers to methods that represent knowledge explicitly using symbols, logical rules and algebraic manipulations. This is the paradigm of classical 'expert systems', where reasoning is performed through structured knowledge bases and inference engines. This is in contrast with connectionist approaches such as neural networks, which learn representations from data but generally lack interpretability and explicit reasoning. Nevertheless, after the second AI winter and the depreciation of AI expert systems, symbolic AI has been marginalised in research for a long time (Kautz 2022). Today, it emerges again, especially in physics because of its closer connections with mathematics, both manipulating symbols with rules (Garnelo & Shanahan 2019).

The main approach dealing with symbolic AI consists in building combination of symbols, leading to extremely large trees that suffer from combinatorial explosion. In physics, Tenachi, Ibata & Diakogiannis (2023) employed units and other known constraints to limit the tree search, for example, rediscovering the various formulae associated with kinetic energy, including the relativist energy of a particle. To facilitate the extraction of relevant symbol chains, gene expression programming (GEP) has been introduced by Ferreira (2001). Trees of varying size and shapes are encoded into genes of fixed size. Optimisation is performed through evolutionary algorithms (Mitchell 1996). Recently, Dominique *et al.* (2021) and Reissmann *et al.* (2021) have employed GEP to discover symbolic expression of the wall pressure spectra Φ_{pp} based on an experimental dataset of a flat plate boundary layer. While GEP provides symbolic expression consistent with existing formulations for zero pressure gradient, they fail to provide unique and readable expressions, limiting the use of symbolic AI towards explainability. In 2023, Lav *et al.* (2023) combined tensor basis neural networks (TBNNs) and GEP to extract interpretable Reynolds stress closures. Note that other alternatives to GEP exist, such as Cartesian gene programming (CGP), for instance, used by Najar *et al.* (2023) to improve shoreline forecasting model. Similarly, Cranmer *et al.* (2020*b*) extended symbolic regression to high-dimensional input such as graphs by using a neural network as an intermediate model. Lemos *et al.* (2023) further applied this technique by first training a graph network while promoting sparse latent representations, and then applying the PySR library to extract explicit physical relations. They show that this coupled approach, encompassing both PoE-iii and PoE-iv, allows AI algorithms to rediscover the laws of orbital mechanics, serving as a validation for future applications where laws are not or partially known.

Note that even with symbolic regression, regularisation is still required to force a sparse representation, potentially leading to explainability (PoE-iv), revealing the close connections between these two pillars of effectiveness. For instance, Kim *et al.* (2021) used an equation learner (EQL) with regularisation. The EQL consists in fully connected networks with a large choice of nonlinear activation functions, acting as a dictionary of possible expression. As a regular network, weights can be learnt from data since the whole symbolic network is end-to-end differentiable. They demonstrate the potential of such a deep network coupled with symbolic regression on mathematics from MNIST images as well as to learn the equation to propagate a dynamical system. It reveals that such hybridisation yields better generalisation capabilities than conventional approaches. They also adapt the method for discovering parametric equations (Zhang *et al.* 2023). In that

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

view, the methodology resembles to the well-known SINDy method proposed by Brunton *et al.* (2016*a*,*b*), where a encoder–decoder architecture is employed, and a classical linear regression on the latent space is achieved with a dictionary of possible functions (figure 8).

I believe that hybridising symbolic regression with deep networks is still not mature today, but deserves a better understanding and testing by the community, either to build robust and explainable surrogate models, or to derive certified or interpretable control laws produced by deep reinforcement learning.

4.5. *Credibility: towards trustworthy AI in CFD*

The previous discussion has focused on the four epistemological pillars of effectiveness for AI in physics. Yet, another pillar, credibility, must be addressed explicitly. In mathematics, the search for truth is tied to rigour and proof, making credibility a natural by-product of the discipline. Algorithms based directly on mathematical principles, such as numerical solvers and mathematical models used in CFD, inherit this trust since their behaviour can be analysed, verified and guaranteed. In contrast, AI models in CFD do not share the same foundation. Their credibility cannot be taken for granted, as they rely heavily on data-driven approximations, stochastic training procedures and high-dimensional parameter spaces that often obscure interpretability. An illustrative example is the issue of non-reproducibility in stochastic training, where neural networks trained under identical conditions can, nevertheless, exhibit significant different behaviours at inference (Pinto, Alguacil & Bauerheim 2022). For this reason, trustworthiness becomes a central concern, not only for ensuring correctness in a strict scientific sense, but also for enabling the adoption of AI methods by a broader CFD community.

As a consequence, achieving credibility in AI for CFD requires more than making models effective: it calls for dedicated algorithms and methodological adaptations that guarantee physical consistency and robustness, which are principles lying beyond our previous PoEs. Addressing these issues is essential for moving beyond proof-of-concept demonstrations of AI in CFD, towards tools that practitioners in fluid mechanics can rely on with the same confidence as in traditional numerical methods. To foster credibility, two specific challenges emerge: (i) robustness and (ii) uncertainty quantification. First, the lack of robustness is often attributed to the use of limited or biased datasets. Unlike numerical solvers, which can be systematically refined through mesh adaptation and error analysis, AI models are only as reliable as the data that shaped them. This is often coined as ‘garbage in, garbage out’. This makes dataset selection and iterative improvement crucial to building models that generalise across the wide variety of flow conditions encountered in practice (§ 4.5.1). Second, even with carefully curated data, AI cannot claim the same rigour and determinism as mathematical proofs or numerical analysis. Nevertheless, credibility can be strengthened by introducing a notion of doubt, i.e. by explicitly quantifying the uncertainties that arise both from data (aleatoric uncertainty) and from the model itself (epistemic uncertainty). Such approaches are explored in § 4.5.2.

4.5.1. *From limited data to robust models*

Robustness in AI for CFD is constrained by the quality and diversity of the training data. Indeed, credibility cannot be achieved if a model is trained on biased or poorly representative datasets. This challenge is even more pronounced in CFD, where each sample often requires significant computational resources, making exhaustive coverage of the parameter space impractical. As a result, improving credibility through robust models must be directly linked to strategies that prioritise the collection of the most relevant

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

and informative samples. An effective solution lies in the development of automatic re-sampling strategies, which iteratively select the most informative samples to maximise metrics such as space coverage or surrogate model performance. By prioritising high-value data points, these strategies significantly reduce the number of samples needed to train a model.

In fluid dynamics, this approach can guide high-fidelity simulations or experiments, focusing computational resources where they are most impactful. Such methods have been successfully applied in traditional machine learning for fluid dynamics, where re-sampling strategies have been combined with Gaussian processes (GPs) for surrogate modelling and optimisation (Roy *et al.* 2018; Campet *et al.* 2020). Another complementary approach involves low-dimensionality reduction techniques, such as principal component analysis (PCA), POD or active subspaces (Constantine, Dow & Wang 2014; Bauerheim *et al.* 2016*b*). These methods identify specific directions in the high-dimensional input space, enabling targeted sampling and efficient exploration. By reducing the complexity of the input domain, they help to guide sampling efforts towards regions that are most relevant for the problem at hand, thereby improving model training while minimising computational costs. These strategies not only enhance surrogate model accuracy, but also pave the way for more effective data-driven approaches in CFD, offering a scalable path for high-dimensional and computationally intensive applications.

Similarly, in deep learning, sampling strategies play a vital role in identifying regions of the parameter space that are uncertain or under-explored. By focusing learning efforts on these areas, models can prioritise data that has the greatest potential to improve performance. This approach, known as active learning, is particularly effective in optimising space coverage or refining solutions along a Pareto front (figure 9). A central ingredient of active learning is the acquisition function $\mathcal{F}$, which assigns a score to each candidate sample x, reflecting its expected usefulness for improving the model. New samples x are then drawn from regions where $\mathcal{F}(x)$ is high, while the network parameters are trained to minimise their predictive loss. The design of the acquisition function is flexible and can integrate multiple criteria relevant to CFD (figure 9v):

$$\mathcal{F} = \mathcal{F}\left(\epsilon(x), \sigma(x), nov(x), \mathcal{O}(x), -\mathcal{C}(x)\right), \tag{4.4}$$

where $\epsilon(x)$ is the predicted error, $\sigma(x)$ the model uncertainty, $\mathcal{O}(x)$ determines the alignment with an optimisation objective and $\mathcal{C}(x)$ evaluates the computational cost in multi-fidelity settings (§ 4.2.4). More recent approaches have further enriched acquisition functions with notions of novelty, diversity and even intrinsic motivation (nov), which encourage exploration of under-represented regions of the parameter space (Michel *et al.* 2025).

For multi-objective optimisation problems, active learning can be implemented using straightforward techniques, such as simply resampling the predicted Pareto front: selection is only guided by the predicted objective, i.e. $\mathcal{F}(\mathcal{O})$, to gather additional data until convergence is achieved. For example, Corban *et al.* (2023) successfully applied this approach to discover optimal wing kinematics using deep learning (figure 9*a*). This method has proven effective in revealing entirely novel solutions. Testing data (• in figures 9*a*(iv) and 9*b*), where kinematics were sampled randomly, illustrate that random datasets often fail to capture optimal configurations, especially regarding C_L, the lift coefficient (figure 9*b*, step 1, where maximum $C_L \approx 0.9$). In contrast, active learning identified iteratively kinematics capable of producing a maximum lift coefficient close to 1.4, significantly higher than the $C_L \approx 0.9$ found in both training and testing datasets based on random sampling (figure 9*b*, step 3). This demonstrates that active learning can discover novel solutions outside the scope of the original training data.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

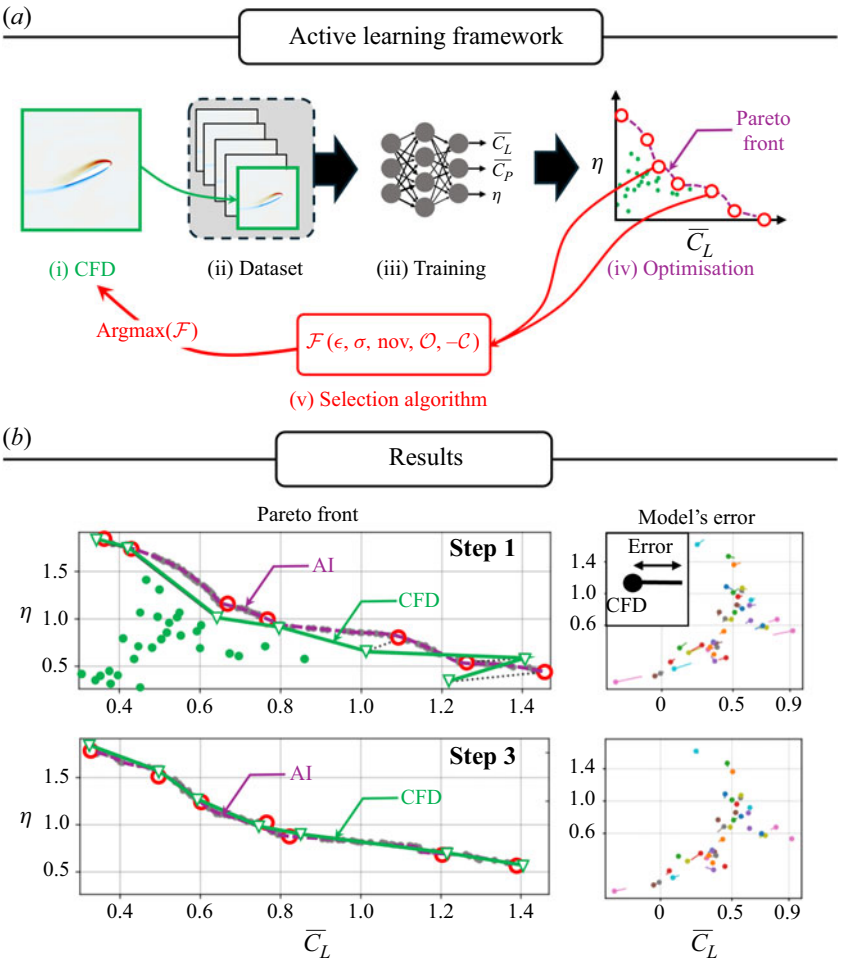


Figure 9. Illustration of the active learning approach introduced by Corban, Bauerheim & Jardin (2023). (*a*) Sketch of the framework where data from CFD (i) are collected in an iterative dataset (ii) to train a deep neural network (iii) predicting the lift (C_L), power (C_P) and efficiency ($\eta = C_L/C_P$) of a 3-D flapping wing depending on its kinematics. NSGA II reuses this network to optimise the kinematics (iv), obtaining a predicted Pareto front (magenta). To guarantee its accuracy, new data are collected along this Pareto front (○) by an acquisition function $\mathcal{F}$ (v), and evaluated by CFD, providing new training data (△). The Pareto front obtained at the first (top) and third (bottom) steps of this iterative active learning are displayed (panel *b*), showing a good convergence. At each step, suboptimal test data are also displayed (right), showing the mismatch between the CFD (●) and network prediction (line).

However, a key limitation of this approach is its lack of full exploration across the input space. This underscores the need to address the exploration–exploitation dilemma inherent in active learning strategies. Recent advances in learning theory offer promising routes to tackle this challenge. For instance, Bobbia & Picard (2024) proposed mathematically explicit ratios to balance exploration (searching broadly across the parameter space) and exploitation (focusing on known high-performance regions). In addition, exploration is closely tied to the notions of novelty and diversity (Michel *et al.* 2025). Instead of measuring distances directly in the input or output domains, where differences may be dominated by noise or superficial variations, recent approaches evaluate distances in feature or latent spaces. These learned representations capture the underlying physical

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

structures and task-relevant patterns, enabling a more meaningful selection of candidate samples for exploration. Applying such strategies to realistic flow scenarios, such as those studied by Corban *et al.* (2023), could further enhance the robustness and discovery potential of active learning frameworks.

While the current implementation has shown promising results in optimisation tasks, active learning can also leverage model's uncertainty (σ in (4.4)), similar to classical sampling strategies employed with Gaussian processes. However, predicting uncertainties in deep networks is far from trivial. Although AI models can achieve impressive predictive accuracy, quantifying uncertainty remains a significant challenge, particularly for high-dimensional systems like fluid flows. A recent review by Koumoutsakos (2024) highlights the need for AI models not only to provide point estimates, but also to offer reliable uncertainty bounds.

4.5.2. *Learning to doubt: quantifying aleatoric and epistemic uncertainties*

Uncertainties in predictions typically arise from two distinct sources: (i) epistemic uncertainty, which stems from the incomplete understanding of the underlying process being modelled; and (ii) aleatoric uncertainty, which refers to the inherent randomness or noise in the data itself, arising from stochastic processes or factors that are intrinsically unpredictable. While Bayesian approaches like Gaussian processes naturally predict the probability density function of the output variable, deep networks need additional ingredients to predict similar quantities.

On one hand, aleatoric uncertainties can be incorporated by assuming that this uncertainty follows a normal distribution $p(y|x) = \mathcal{N}(\mu, \sigma^2)$. Mean and variance can then be estimated by a neural network outputting both quantities (μ_θ and σ_θ^2). To train the model, a negative log-likelihood (NLL) loss function can be employed, which for a Gaussian distribution with estimated mean and variance becomes

$$\mathcal{L}_{NLL}(y, \mu_\theta(x), \sigma_\theta^2(x)) = \frac{1}{2}\log\sigma_\theta^2(x) + \frac{(y-\mu_\theta(x))^2}{2\sigma_\theta^2}. \tag{4.5}$$

Alternatively, a β-NLL loss function can be employed during training, as proposed by Seitzler *et al.* (2022), which allows for better calibration of uncertainty estimates, particularly in scenarios where the model might be overconfident, as well as out-of-distribution scenarios.

On the other hand, methods such as deep ensemble (Lakshminarayanan, Pritzel & Blundell 2017, DE) can be employed to model epistemic uncertainty, which arises from the model's lack of knowledge about the underlying process. This is achieved by training multiple instances of the same model with different random initialisations (figure 10*a*, left). The resulting uncorrelated models can be used to estimate both the mean ($\mathbb{E}(y)$) and variance ($\mathbb{V}(y)$) of the predicted outputs. However, the significant drawback of this approach is the substantial computational cost due to the need for multiple training runs and inferences. An alternative strategy is to generate multiple models from a single training process (figure 10*a*, right), such as using dropout techniques (Gal & Ghahramani 2016). This approach drastically reduces computational cost by leveraging the stochasticity inherent in dropout layers, which randomly deactivate neurons during training. However, while this reduces the computational burden, it also leads to highly correlated models, which can compromise the quality of the uncertainty estimates, particularly the variance predictions. Other promising alternatives include deep Bayesian networks, which explicitly model uncertainty by placing a prior distribution over the

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

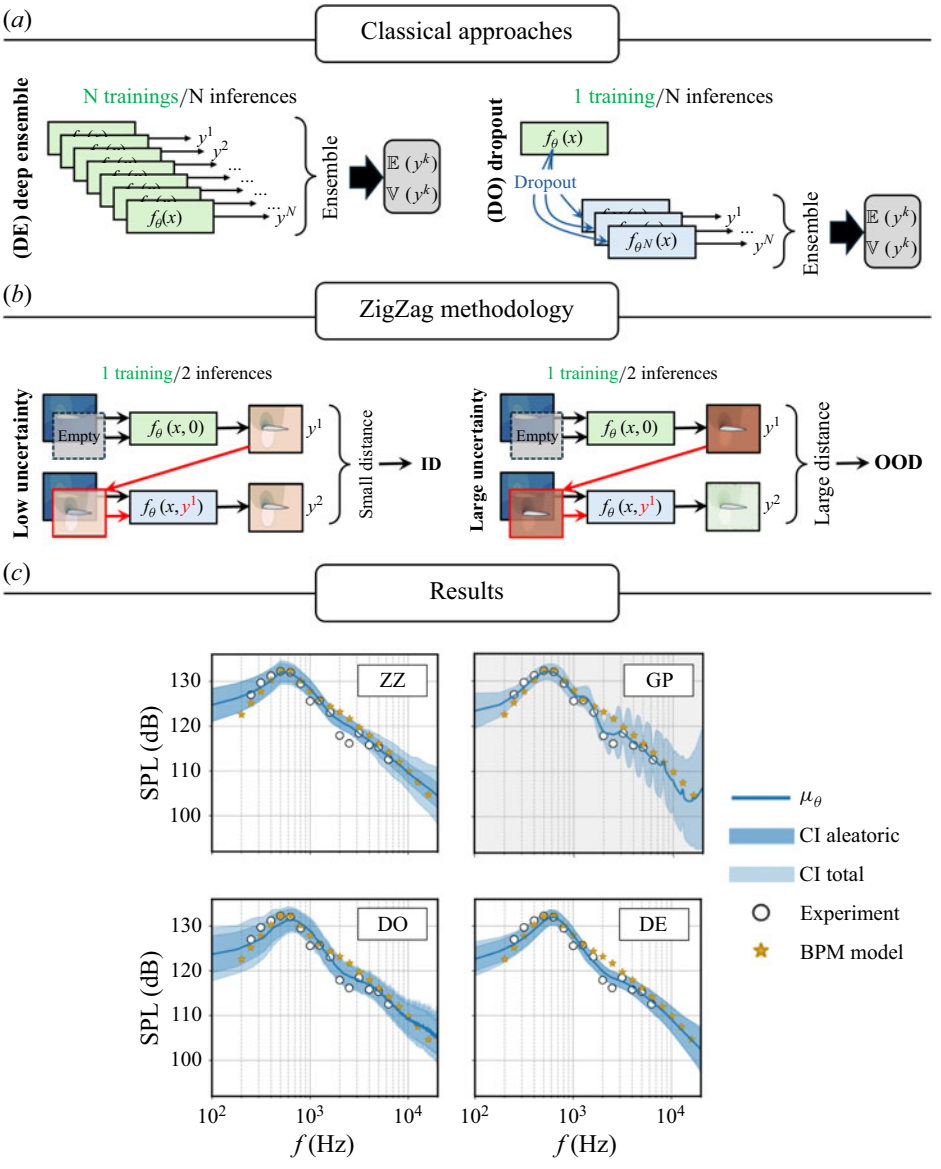


Figure 10. (*a*) Illustration of classical approaches for uncertainty estimations through a variance $\mathbb{V}(y)$: deep ensemble (DE, left) and dropout (DO, right). (*b*) Sketch of the zigzag (ZZ) methodology introduced by Durasov *et al.* (2024). It consists in a network $f_\theta(x, y)$ which is fed inputs x, as well as by its own prediction y. A two step approach, starting from $y = 0$, provides a first answer y^1, which is then fed for a second step, providing a new output y^2. The distance $\|y^2 - y^1\|$ measures the uncertainty of the model, and thus allows to discriminate between ID or OOD scenario. (*c*) Uncertainty estimations by deep ensemble (DE), Monte Carlo dropout (DO), zigzag (ZZ) and Gaussian process (GP) applied to the trailing edge noise dataset (Brooks, Pope & Marcolini 1989; Markelle, Longjohn & Nottingham 2014). It shows the mean prediction μ_θ, as well as the aleatoric and total uncertainty (confidence intervals, CI).

model's parameters. However, these networks are more difficult to train effectively and are computationally demanding (Mackay 1995).

Recently, new indirect approaches have emerged, such as zigzag (see figure 10*b*) introduced by Durasov *et al.* (2024). This method modifies a standard neural network

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

$y = f_\theta(x)$ into a reentrant network where the output $y^{k+1} = f_\theta(x, y^k)$ is iteratively refined through multiple stages. The first iteration y^0 is typically initialised to 0. For successful training, the network must satisfy $f_\theta(x, 0) \approx f_\theta(x, y) \approx y$ for all training pairs (x, y). Recent theoretical results (Foglia *et al.* 2025) show that the difference $\|y^2 - y^1\|$ serves as an effective proxy for total uncertainty. Compared with dropout (DO), deep ensemble (DE) or Gaussian process (GP), this method provides accurate uncertainty estimations while reducing the training and inference costs: only one training and two inferences (figure 10*c*). The idea behind the method is to evaluate if the network changes its answer when fed by its own prediction, as a human would do when 'thinking twice'. In addition, when for the same input x, several outputs y are available (e.g. repeated experiment, CFD with different models), aleatoric and epistemic uncertainties can be distinguished.

Nevertheless, while these methods show promise, their evaluation has mostly been limited to toy problems. More comprehensive benchmarks, particularly on real-world high-dimensional fluid dynamics problems, are needed to validate their effectiveness in practical applications. A notable example of such current evaluation is the quantification of uncertainties in auto-encoders for 2-D aerodynamic predictions (Saetta, Tognaccini & Iaccarino 2024). Another example is provided by Alguacil *et al.* (2024), where active learning is combined with UQ estimations to build sample-efficient network models of the acoustic cascade response, which involves singularities difficult to capture by random sampling.

In addition, identifying out-of-distribution (OOD) data is critical for evaluating the confidence of a model's predictions. OOD refers to data that the model has not encountered during training. Since testing pairs rarely match exactly with those observed during training, the goal is to identify features that are either unique or significantly different from the initial training data. OOD data can be assessed indirectly by predicting uncertainties, as OOD inputs are more likely to yield uncertain predictions. In this context, zigzag can again be used, as it distinguishes between in-distribution (ID) and OOD data: for ID data, $y^2 \approx y^1$ (as shown in figure 10*b*, left), whereas for OOD data, $y^2 \neq y^1$ (figure 10*b*, right). Further development and evaluation of OOD detection methods are needed in future research. This is especially relevant in optimisation pipelines that employ deep neural networks, as optimal solutions are often OOD never seen during training.

4.6. *Beyond the four pillars: generative and foundation models*

Although not directly tied to our four PoEs, generative models (§ 4.6.1) and self-supervised foundation models (§ 4.6.2) merit brief discussion due to their transformative potential.

4.6.1. *Generative models*

In contrast with surrogate networks that learn a deterministic mapping $y = f(x)$, generative models aim instead to approximate the full conditional distribution $p(y|x)$, from which new consistent samples $y^* \sim p(y|x)$ can be drawn. The remarkable success of generative AI in areas like natural language processing (Brown *et al.* 2020; Touvron *et al.* 2023) and computer vision (Yuan *et al.* 2021) has inspired growing interest in their applications to physics, particularly in fluid dynamics. Generative models, such as generative adversarial networks (GANs), variational autoencoders (VAEs) and diffusion models (DMs), offer innovative approaches (see table 2) to create high-fidelity fluid simulations by learning the underlying data distributions $p(y|x)$. These models provide a compelling alternative to traditional simulation methods, enabling

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

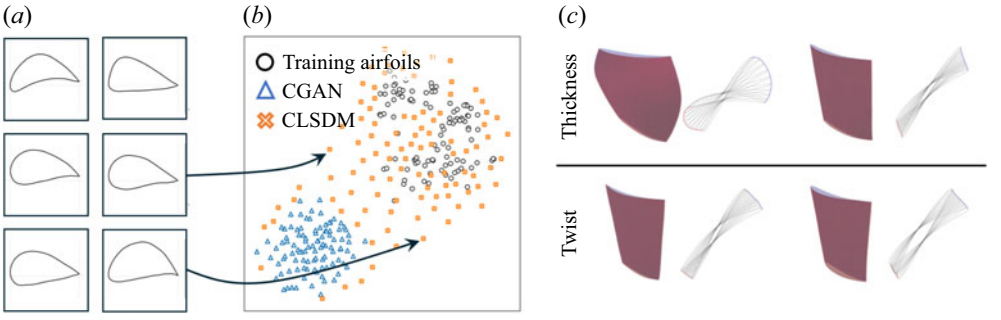


Figure 11. (*a*) Results obtained by Wei *et al.* (2024) with a conditioned DM (CLSDM) to generate possible aerofoil geometries. (*b*) Visualisation of the results using a t-sne map revealed that CLSDM can generate more diverse blade better than CGAN, especially when only limited training data are available (here, 100 training aerofoils). (*c*) Example of 3-D turbomachinery blades and their 2-D projection generated by CLSDM, when constrained on thickness (top) and twist (bottom).

the generation of new flow fields or turbulence patterns. For instance, Xie *et al.* (2018) employed a temporally consistent GAN to achieve super-resolution fluid flows for movie animations. Similarly, Drygala *et al.* (2022) demonstrated that deep convolutional GANs (DCGANs) outperformed classical GANs and Wasserstein GANs (WGANs) in generating turbulent fields in a blade cascade configuration. Further advancements in this domain include Li *et al.* (2023), who showcased the effectiveness of CNNs and GANs in inferring velocity components in rotating turbulent flows. More recently, Shu, Li & Farimani (2023) used a physics-guided denoising diffusion probabilistic model (DDPM) to generate turbulence fields. Building on this foundation, Qiu *et al.* (2024) extended the approach to generate space–time fluid flows, demonstrating its applicability to complex scenarios such as the Von Karman vortex street and 3-D brain artery flows.

Similarly, Wei *et al.* (2024) explored the generation of new geometries, such as 2-D aerofoils and 3-D turbomachinery blades, using a conditional latent space diffusion model (CLSDM, figure 11). The conditioning of the model, which guides sample generation with geometrical constraints, was enforced during the denoising steps. Results demonstrated the exceptional efficiency of CLSDM in generating both diverse and novel geometries, even with a limited number of training examples (e.g. only 100 shapes for panels *a* and *b*). A comparative analysis with other techniques, such as CGANs, highlights the superiority of CLSDM. For example, t-SNE mapping (figure 11*b*) showed that CLSDM generates a broader and more novel range of shapes compared with CGANs, particularly when training data are limited (Wei *et al.* 2024). These generated shapes were particularly effective in building training datasets for surrogate modelling in shape optimisation tasks, such as predicting and optimising the lift-to-drag ratio (L/D). Similarly, CLSDM is able to generate realistic 3-D blades (figure 11*c*), with physical constraints as thickness (top) or twist distribution (bottom). These generated blades have been integrated in a simplified pre-design framework to optimise the isentropic efficiency of a turbomachinery, revealing both their good performance and their adherence to physical constraints. Thus, while diffusion models are still recent and underexplored in fluid mechanics, their potential merits deeper investigation. For example, turbulence is inherently stochastic, yet most simulation tools are deterministic. In contrast, generative models can capture the full data distribution, thereby extending beyond the classical deterministic paradigm of CFD. Future research should focus on enhancing our understanding of these methods and applying it to more realistic configurations and scenarios.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

4.6.2. *Foundation models*

Foundation models consist in large pre-trained models that can be fine-tuned for specific tasks. They offer significant potential for applications in physics (Subramanian *et al.* 2023), particularly in fluid mechanics. These models provide a versatile and powerful framework for tackling a wide range of fluid dynamics problems, from aerodynamics to weather forecasting, serving as valuable tools for both researchers and engineers. Unlike traditional approaches that incorporate semantics (PE-iv) through inductive biases in deep learning networks, foundation models often rely on transformer architectures. Their key principle is to eliminate the constraints imposed by inductive biases, instead leveraging vast amounts of data to pre-train networks that reconstruct partially masked inputs (in-painting). This enables the multi-head attention mechanism of transformers to learn contextualised, meaningful embeddings relevant to the data. However, as Jakubik *et al.* (2023) cautioned, 'it would be unwise to anticipate superior performance (of foundation models) across all application categories compared with state-of-the-art models'.

Nonetheless, foundation models show clear advantages in scenarios with limited labelled data, a strength also shared with other pre-trained and fine-tuning strategies discussed in § 4.2.4. Achieving the full potential of foundation models presents new challenges beyond our pillars of effectiveness, particularly in providing easy and fast access to large and diverse datasets. Initiatives like the Johns Hopkins Turbulence Database (JHTDB), BlastNet and, more recently, The Well (Ohana *et al.* 2024), aim to facilitate access to datasets relevant for physics and fluid mechanics. Additionally, this effort necessitates the harmonisation of database formats, variables and structures, as well as the development of open-source pre-processing tools to efficiently manipulate these datasets. First demonstrations of foundation models have been realised recently by Liu *et al.* (2024), Tierz *et al.* (2024) and finally Feng *et al.* (2025), who proposed FluidZero, a pre-trained generative model able to address diverse tasks such as predictions, shape optimisation, parameter identification or control problems. This shows the potential of versatile foundation models in fluid mechanics, opening possibly new path to solve fluid problems in the near future.

4.6.3. *Rethinking effectiveness for generative and foundation models*

It is important to emphasise that generative and foundation models do not necessarily incorporate the four aspects of mathematical effectiveness discussed in this article. Indeed, they both represent a distinct paradigm within AI that differs significantly from traditional physics-inspired modelling and CFD. This raises the question of whether the four pillars of effectiveness can also be applied to such models. In principle, certain pillars already find partial implementation. For instance, scale separation is central to the architecture of UNets, which are widely used in diffusion models or foundation models. Similarly, symmetries can be enforced in generative frameworks through equivariant architectures or data augmentation (Xie *et al.* 2018). However, other pillars appear to contradict the foundational philosophy of these models. Sparsity, for example, is opposed to the overparametrisation that underlies their success and semantics is never explicitly encoded, but rather should emerge implicitly, if at all, from large-scale training. This suggests that while generative and foundation models can accommodate some of these principles, their integration is selective and often indirect, highlighting a fundamental difference between models designed around effectiveness and explicability, and those designed around universality and scale.

Consequently, this raises a critical question: are these approaches suitable for physics? Relying on such vast datasets, which are often challenging to harmonise, store, validate

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

and share, may not always be practical, especially when compared with simpler multi-fidelity pipelines tailored for specific tasks. For example, while foundation models in fields like natural language processing are justified by the requirement to capture long dependencies and intricate contexts, the case for their use in physics is less clear. Physical phenomena are often characterised by spatial or temporal locality, and governed by well-defined equations that provide inherent structure and constraints, such as conservation laws and boundary conditions. In addition, training foundation models requires immense computational resources and energy, which may not always justify the gains in accuracy or flexibility, particularly when simpler, task-specific models can achieve comparable performance. Note however that this conclusion might be different with experimental data subject to noise and experimental biases.

Nevertheless, if generative and foundation models indeed represent a paradigm shift in CFD, this suggests that the present framework of epistemological pillars for effectiveness may need to be revisited in the near future. In particular, pillars derived from mathematics in the context of physics might not fully capture the requirements of these new approaches. For these models, statistics for example, an aspect of mathematics not scrutinised in this work, or alternative conceptual references, could provide a more suitable foundation for defining effectiveness. As an illustration, while mathematical logic in this work has motivated the principle of rules with semantics and explainability, future perspectives might reconsider this principle through the lens of causality going beyond mere correlations, which is more aligned with the statistical paradigm behind generative models. Exploring such alternative formulations constitutes an important direction for future research, opening the way towards new epistemological pillars adapted to these specific AI paradigms that emphasise universality, scale and stochasticity.

5. Conclusion

AI for CFD is a rapidly evolving field that holds the potential to revolutionise how we approach problems in fluid dynamics. The integration of AI into CFD not only promises to accelerate simulations and improve accuracy, but also opens new pathways for understanding and discovering novel fluid phenomena. It also opens the path to bridge the gap between simulations and experiments by extracting relevant features from the best of these two worlds. To make these new AI approaches feasible and effective on complex problems, I proposed four pillars of effectiveness (PoEs) by examining why mathematics is so effective in physics, namely: (PoE-i) symmetries providing internal structures to equations; (PoE-ii) scales separation allowing a specific treatment of the physics for each scale and their interactions; (PoE-iii) sparsity, which promotes explicability; and finally (PoE-iv) semantic significance, which gives mathematics not a purely grammatical structure, but also a meaning, facilitating abstraction and reasoning. Beyond these PoEs, AI also suffers from lack of guarantees, compared with the rigour of mathematics. This suggests consideration of a fifth pillar, credibility, which is envisioned as a pivotal role not only to improve effectiveness, but also to foster a wide adoption of AI in the scientific community. I then discuss if, and how, these pillars can be applied to AI algorithms in CFD. The conclusion is that those pillars are actually declined in a succession of technical advances that have shown promising results when using AI in CFD. These results highlight the potential of this emerging technology to be, possibly, as effective as standard numerical tools. We can even envision that in the near future, classical simulations and artificial intelligence will have increasingly more blurry boundaries, where the two approaches will be merged into hybridised numerical tools. While AI techniques will undoubtedly evolve over time, with some learning algorithms emerging,

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

while others are falling into disuse, I firmly believe that the methodologies grounded in the four PoEs will endure. These foundational principles aim to provide a robust framework for understanding and improving AI effectiveness in CFD and, more generally, to computational science. Nevertheless, generative and foundation models may require a reformulation of these PoEs in the near future, since they introduce a paradigm shift that departs significantly from the surrogate modelling and CFD paradigms addressed in this work.

Acknowledgements. I thank B. Dubrulle and D. Barkley for their enlightening discussions to improve this perspective.

Funding. This work has benefited from the AI Interdisciplinary Institute ANITI. ANITI is funded by the France 2030 program under the Grant agreement no. ANR-23-IACL-0002.

Declaration of interests. The author reports no conflict of interest.

References

Aguilar, J.G., Dawson, J.R., Schuller, T., Durox, D., Prieur, K. & Candel, S. 2021 Locking of azimuthal modes by breaking the symmetry in annular combustors. *Combust. Flame* **234**, 111639.

Ajuria-Illaramendi, A., Bauerheim, M. & Cuenot, B. 2022 Performance and accuracy assessments of an incompressible fluid solver coupled with a deep convolutional neural network. *Data-Centric Engng* **3**, e2.

Alguacil, A., Bauerheim, M., Jacob, M.C. & Moreau, S. 2022 Deep learning surrogate for the temporal propagation and scattering of acoustic waves. *AIAA J.* **60** (10), 5890–5906.

Alguacil, A., Bauerheim, M., Moreau, S., Lavoie, P. & Bolduc-Teasdale, F. 2024 Data-driven acceleration and sensitivity study of an acoustic cascade response model. In *AIAA Aviation Forum*, p. 3389.

Alvarado, R. 2023 Ai as an epsitemic technology. *Sci. Engng Ethics* **29**, 32.

Anderson, P.W. 1972 More is different, broken symmetry and the nature of the hierarchical structure of science. *Science* **177** (4047), 393–396.

Anderson, P.W. & Abrahams, E. 2009 Machines fall short of revolutionary science. *Science* **324**, 1515–1516.

Azar, A.T., *et al.* 2021 Drone deep reinforcement learning: a review. *Electronics* **10** (9), 999.

Bachelard, G. 1934 *Le Nouvel Esprit Scientifique*. Librairie Félix Alcan.

Bachelard, G. 1936 *L'expérience de l'espace dans la physique contemporaine*. PUF.

Bachelard, G. 1938 *La formation de l'esprit scientifique*. Librairie Philosophique J. Vrin.

Baddoo, P.J., Herrmann, B., McKeon, B.J., Kutz, N. & Brunton, S.L. 2023 Phyiscs-informed dynamic mode decomposition. *Proc. R. Soc. Lond. A* **479**, 20220576.

Bar-Sinai, Y., Hoyer, S., Hickey, J. & Brenner, M.P. 2019 Learning data-driven discretizations for partial differential equations. *Proc. Natl Acad. Sci. USA* **116** (31), 15344–15349.

Barbara, N.H., Wang, R. & Machester, I.R. 2024 On robust reinforcement learning with Lipschitz-bounded policy networks. arXiv 2405.11432v1.

Barbiero, P., Ciravegna, G., Giannini, G., Lio, P., Gori, M. & Melacci, S. 2022 Entropy-based logic explanations of neural networks. In *AAAI Conference on Artificial Intelligence*, vol. 36.

Bassi, F. & Rebay, S. 1997 A high-order accurate discontinuous finite element method for the numerical solution of the compresible Navier–Stokes equations. *J. Comput. Phys.* **131** (2), 267–279.

Bauerheim, M., Cazalens, M. & Poinsot, T. 2014*a* A theoretical study of mean azimuthal flow and asymmetry effects on thermo-acoustic modes in annular combustors. *Proc. Combust. Inst.* **35** (3), 3219–3227.

Bauerheim, M., Ndiaye, A., Constantine, P., Moreau, S. & Nicoud, F. 2016*a* Symmetry breaking of azimuthal thermoacoustic modes: the UQ perspective. *J. Fluid Mech.* **789**, 534–566.

Bauerheim, M., Nicoud, F. & Poinsot, T. 2016*b* Progress in analytical methods to predict and control azimuthal combustion instability modes in annular chambers. *Phys. Fluids* **28** (2), 021303.

Bauerheim, M., Salas, P., Nicoud, F. & Poinsot, T. 2014*b* Symmetry breaking of azimuthal thermoacoustic modes in annular cavities: a theoretical study. *J. Fluid Mech.* **760**, 431–465.

Beck, A., Zeifang, J., Schwarz, A. & Flad, D.G. 2020 A neural network based shock detection and localization approach for discontinuous Galerkin methods. *J. Comput. Phys.* **423**, 109824.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

BERGER, M.J. & OLIGER, J. 1984 Adaptive mesh refinement for hyperbolic partial differential equations. *J. Comput. Phys.* **53** (3), 484–512.
BERGER, S., ARROYO-RAMO, A., GUILLET, V., LAHIRE, T., MARTIN, B., JARDIN, T., RACHELSON, E. & BAUERHEIM, M. 2024 Reliability assessment of off-policy deep reinforcement learning: a benchmark for aerodynamics. *Data-Centric Engng* **5**, e2.
BÉTHUNE, L., BOISSIN, T., SERRURIER, M., MAMALET, F., FRIEDRICH, C. & GONZALEZ-SANZ, A. 2022 Pay attention to your loss: understanding misconceptions about 1-Lipschitz neural networks. In *NeurIPS*, vol. 36.
BOBBIA, B. & PICARD, M. 2024 Active learning for regression based on wasserstein distance and groupsort neural networks. arXiv 2403.15108.
BOHR, N. 1958 *Atomic Physics and Human Knowledge*. Chapman & Hall.
BOMMASANI, R., *et al.* 2022 On the opportunities and risks of foundation models. arXiv 2108.07258.
BOZINOVSKI, S. & FULGOSI, A. 1976 The influence of pattern similarity and transfer learning upon training of a base perceptron B2. *Proc. Symp. Informatica* **3** (121), 5.
BRONSTEIN, M.M., BRUNA, J., COHEN, T. & VELICKOVIC, P. 2021 Geometric deep learning: grids, groups, graphs, geodesic and gauges. arXiv 2104.13478v2.
BROOKS, T.F., POPE, D.S. & MARCOLINI, M. 1989 Airfoil self-noise and prediction. *Tech. Rep*. NASA.
BROWN, T., *et al.* 2020 Language models are few-shot learners. In *Advances in Neural Information Processing Systems* (ed. H. LAROCHELLE, M. RANZATO, R. HADSELL, M.F. BALCAN & H. LIN), vol. 33, pp. 1877–1901. Curran Associates, Inc.
BRUNTON, S.L., PROCTOR, J.L. & KUTZ, J.N. 2016*a* Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl Acad. Sci. USA* **113** (15), 3932–3937.
BRUNTON, S.L., PROCTOR, J.L. & KUTZ, J.N. 2016*b* Sparse identification of nonlinear dynamics with control (SINDYc). *IFAC-PapersOnLine* **49** (18), 710–715.
BUHENDWA, A.B., BEZGIN, D.A. & ADAMS, N.A. 2022 Consistent and symmetry preserving data-driven interface reconstruction for the level-set method. *J. Comput. Phys.* **457**, 111049.
CAI, S., MAO, Z., WANG, Z., YIN, M. & KARNIADAKIS, G.E. 2021 Physics-informed neural networks (PINNs) for fluid mechanics: a review. *Acta Mechanica Sin.* **37**, 1727–1738.
CAMPET, R., ROY, P.T., CUENOT, B., RIBER, E. & JOUHAUD, J.C. 2020 Design optimization of an heat exchanger using Gaussian process. *Intl J. Heat Mass Transfer* **150**, 119264.
CANTWELL, B. 1978 Similarity transformations for the two-dimensional, unsteady, stream-function equation. *J. Fluid Mech.* **85**, 257–271.
CATALANI, G., AGARWAL, S., BERTRAND, X., TOST, F., BAUERHEIM, M. & MORLIER, J. 2024 Neural fields for rapid aircraft aerodynamics simulations. *Sci. Rep.* **14** (1), 25496.
CATALANI, G., COSTERO, D., BAUERHEIM, M., ZAMPIERI, L., CHAPIN, V., GOURDAIN, N. & BAQUÉ, P. 2023 A comparative study of learning techniques for the compressible aerodynamics over a transonic RAE2822 airfoil. *Comput. Fluids* **251**, 105759.
CHEN, R.T., RUBANOVA, Y., BETTENCOURT, J. & DUVENAUD, D. 2018 Neural ordinary differential equations. In *NIPS*, vol. 32.
CHENG, L., ILLARRAMENDI, E.A., BOGOPOLSKY, G., BAUERHEIM, M. & CUENOT, B. 2021 Using neural networks to solve the 2D Poisson equation for electric field computation in plasma fluid simulations. arXiv 2109.13076v3.
CHENG, S., MOREL, R., ALLYS, E., MÉNARD, B. & MALLAT, S. 2024 Scattering spectra models for physics. *PNAS* **3** (4), 103.
CHNAFA, C., MENDEZ, S., MORENO, R. & NICOUD, F. 2015 Using image-based CFD to investigate the intracardiac turbulence. In *Modeling the Heart and the Circulatory Systems*, pp. 97–117.
COIFFIER, G. & BÉTHUNE, L. 2024 1-Lipschitz neural distance fields. *Comput. Graph. Forum* **43** (5), e15128.
COLIN, O., DUCROS, F., VEYNANTE, D. & POINSOT, T. 2000 A thickened flame model for large eddy simulations of turbulent premixed combustion. *Phys. Fluids* **12** (7), 1843–1863.
COLOMBO, M., MORLIER, J. & BAUERHEIM, M. 2023 Prediction of gust aeroelastic performance of hale using graph neural networks. In *AIAA Scitech 2023 Forum*, p. 2570.
CONSTANTINE, P.G., DOW, E. & WANG, Q. 2014 Active subspace methods in theory and pratice: applications to Kriging surfaces. *SIAM J. Sci. Comput.* **36** (4), 1500–1524.
CORATGER, T., FARAG, G., ZHAO, S., BOIVIN, P. & SAGAUT, P. 2021 Large-eddy lattice-Boltzmann modeling of transonic flows. *Phys. Fluids* **33** (11), 115112.
CORBAN, B., BAUERHEIM, M. & JARDIN, T. 2023 Discovering optimal flapping wing kinematics using active learning. *J. Fluid Mech.* **974**, A54.
CORRY, L. 1996 *Modern Algebra and the Rise of Mathematical Structures*. Birkhäuser.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

Cranmer, M., Greydanus, S., Hoyer, S., Battaglia, P., Spergel, D. & Ho, S. 2020*a* Lagrangian neural networks. arXiv 2003.04630.
Cranmer, M., Sanchez-Gonzalez, A., Battaglia, P., Xu, R., Cranmer, K., Spergel, D. & Ho, S. 2020*b* Discovering symbolic models from deep learning with inductive biases. In *NeurIPS*, vol. 34.
Curie, P. 1984 Sur la symétrie dans les phénomènes physiques, symétrie d'un champ électrique et d'un champ magnétique. *J. Phys. Théor. Appl.* **3** (1), 393–415.
D'Aniello, R., Gövert, S., Knobloch, K. & Kissner, C. 2023 A general acoustic framework for the assessment of noise emitted by combustors: a first case study. *J. Sound Vib.* **562**, 117846.
Doan, N.A.K., Polifke, W. & Magri, L. 2020 Physics-informed echo state networks. arXiv 2011.02280.
Dominique, J., Christophe, J., Schram, C. & Sandberg, R.D. 2021 Inferring empirical wall pressure spectral models with gene expression programming. *J. Sound Vib.* **506**, 116162.
Doshi-Velez, F. & Kim, B. 2017 Towards a rigorous science of interpretable machine learning. arXiv 1702.08608v2.
Drozda, L., Mohanamuraly, P., Cheng, L., Lapeyre, C., Daviller, G., Realpe, Y., Adler, A., Staffelbach, G. & Poinsot, T. 2023 Learning an optimised stable Taylor–Galerkin convection scheme based on a local spectral model for the numerical error dynamics. *J. Comput. Phys.* **493**, 112430.
Drygala, C., Winhart, B., di Mare, F. & Gottschalk, H. 2022 Generative modeling of turbulence. *Phys. Fluids* **34** (3), 035114.
Duan, W. 2023 The challenge of artificial intelligence scientists to the epistemology of science. *J. Library Inform. Sci. Agric.* **35** (11), 4–12.
Dubrulle, B. & Frisch, U. 1991 Eddy viscosity of parity-invariant flow. *Phys. Rev. A* **43**, 5355.
Durasov, N., Oner, D., Donier, J., Le, H. & Fua, P. 2024 Enabling uncertainty estimation in iterative neural networks. In *ICML* , vol. 41.
Einstein, A. 1920 *Relativity: the Special and General Theory*. Methuen & Co Ltd.
Fakhar, K., Hayat, T., Yi, C. & Zhao, K. 2009 Symmetry transformation of solutions for the Navier–Stokes equations. *Appl. Maths Comput.* **207** (1), 213–224.
Fanaskov, V. 2024 Neural multigrid architectures. arXiv 2402.05563v1.
Farahani, A., Voghoei, S., Rasheed, K. & Arabnia, H.R. 2020 A brief review of domain adaptation. arXiv 2010.03978v1.
Faure-Beaulieu, A., Indlekofer, T., Dawson, J.R. & Noiray, N. 2021 Imperfect symmetry of real annular combustors: beating thermoacoustic modes and heteroclinic orbits. *J. Fluid Mech.* **2021** (649), R1.

Faure-Beaulieu, A. & Noiray, N. 2022 Symmetry breaking of azimuthal waves: slow-flow dynamics on the Bloch sphere. *Phys. Rev. Fluids* **5**, 023201.
Feng, H., Zheng, H., Hu, P., Liu, H., Yu, C., Wei, L., Feng, R., Duan, J., Fan, D. & Wu, T. 2025 Fluidzero: mastering diverse tasks in fluid systems through a single generative model. arXiv (under review).

Ferreira, C. 2001 Gene expression programming: a new adaptive algorithm for solving problems. *Complex Syst.* **13** (2), 87–129.
Florack, L., Romeny, B.M.T.H., Koenderink, J.J. & Viergever, M.A. 1992 Scale and the differential structure of images. *Image Vis. Comput.* **10** (6), 376–388.
Florack, L.M.J., Romeny, B.M.T.H., Koenderink, J.J. & Viergever, M.A. 1994 General intensity transformations and differential invariants. *J. Math. Imag. Vis.* **4**, 171–187.
Foglia, E., Bobbia, B., Durasov, N., Bauerheim, M., Fua, P., Moreau, S. & Jardin, T. 2025 Do you understand epistemic uncertainty? think again! rigorous frequentist epistemic uncertainty estimation in regression. arXiv 2503.13317v1.
Gal, Y. & Ghahramani, Z. 2016 Dropout as a bayesian approximation: representing modeul uncertainty in deep learning. In *ICML*, pp. 1050–1059.
Ganascia, J.-G. 2010 Epistemology of AI revisited in the light of the philosophy of information. *Knowledge Technol. Policy* **23**, 57–73.
Garnelo, M. & Shanahan, M. 2019 Reconciling deep learning with symbolic artificial intelligence: representing objects and relations. *Curr. Opin. Behav. Sci.* **29**, 17–23.
Garnier, P., Viquerat, J., Rabault, J., Larcher, A., Kuhnle, A. & Hachem, E. 2021 A review on deep reinforcement learning for fluid mechanics. *Comput. Fluids* **225**, 104973.
Gens, R. & Domingos, P.M. 2014 Deep symmetry networks. In *NIPS*, vo. 27.
Germano, M., Piomelli, U., Moin, P. & Cabot, W. 1991 A dynamic subgrid-scale eddy viscosity model. *Phys. Fluids* **3** (7), 1760–1765.
Gode, H., Kruse, C., Angersbach, R., Köstler, H., Bauerheim, M. & Rüde, U. 2024 Comparison of multigrid and machine learning-based poisson solvers. In *HAL*, p. 04689025.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

GOLUBITSKY, M. & STEWART, I. 1986 Symmetry and stability in Taylor–Couette flow. *SIAM J. Math. Anal.* **17** (2), 249–288.

GOODFELLOW, I., POUGET-ABADIE, J., MIRZA, M., XU, B., WARDE-FARLEY, D., OZAIR, S., COURVILLE, A. & BENGIO, Y. 2014 Generative adversarial nets. In *Advances in Neural Information Processing Systems*.

GRATTAROLA, D. & VANDERGHEYNST, P. 2022 Generalised implicit neural representations. In *Conference on Neural Information Processing Systems*, vol. 36.

GRAVES, A., MOHAMED, A.-R. & HINTON, G. 2013 Speech recognition with deep recurrent neural networks. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*.

GREBENEV, V.N. & OBERLACK, M. 2007 Approximate Lie symmetries of the Navier–Stokes equations. *J. Nonlinear Math. Phys.* **14** (2), 157–163.

GREYDANUS, S., DZAMBA, M. & YOSINSKI, J. 2019 Hamiltonian neural networks. In *NIPS*, vol. 33.

GRIZZI, S. & CAMUSSI, R. 2012 Wavelet analysis of near-field pressure fluctuations generated by a subsonic jet. *J. Fluid Mech.* **698**, 93–124.

HERNANDEZ, Q., BADIAS, A., CHINESTA, F. & CUETO, E. 2023 Thermodynamics-informed neural networks for physically realistic mixed reality. *Comput. Meth. Appl. Mech. Engng* **407**, 115912.

HERRMANN, L. & KOLLMANNSBERGER, S. 2024 Deep learning in computational mechanics: a review. *Comput. Mech.* **74**, 281–331.

HO, J., JAIN, A. & ABBEEL, P. 2020 Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems* (ed. H. LAROCHELLE, M. RANZATO, R. HADSELL, M.F. BALCAN & H. LIN), vol. 33, pp. 6840–6851. Curran Associates, Inc.

HUSSAIN, M.S., ZAKI, M.J. & SUBRAMANIAN, D. 2022 Global self-attention as a replacement for graph convolution. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, vol. 28.

ILLARRAMENDI, E.A., ALGUACIL, A., BAUERHEIM, M., CUENOT, B. & BENAZERRA, B. 2020*a* Towards an hybrid computational strategy based on deep learning for incompressible flows. In *AIAA Aviation Forum*, p. 3058.

ILLARRAMENDI, E.A., BAUERHEIM, M. & CUENOT, B. 2020*b* Analysis of downscaled branches and receptive field on a cnn-based incompressible solver. In *APS Division Fluid Dynamics Meeting*, p. p20035.

JAKUBIK, J., *et al.* 2023 Foundation models for generalist geospatial artificial intelligence. arXiv 2310.18660v2.

JEBEILE, J. 2024 From regional climate models to usable information. *Climatic Change* **177**, 53.

KAISER, B., WU, T., SONNEWALD, M., THACKRAY, C. & CALLIS, S. 2024 A moonshot for AI oracles in the sciences. arXiv 2406.17836.

KAUTZ, H.A. 2022 The third AI summer: AAAI Robert S. Engelmore memorial lecture. *AI Mag.* **43**, 105–125.

KIM, S., LU, P.Y., MUKHERJEE, S., GILBERT, M., JING, L., CEPERIC, V. & SOLJACIC, M. 2021 Integration of neural network-based symbolic regression in deep learning for scientific discovery. *IEEE Trans. Neural Networks Learn. Syst.* **32** (9), 4166.

KOESTLER, L., GRITTNER, D., MOELLER, M., CREMERS, D. & LÄHNER, Z. 2022 Intrinsic neural fields: learning functions on manifolds. *Comput. Vis. – ECCV* **17**, 622–639.

KOUMOUTSAKOS, P. 2024 On roads less travelled between AI and computational science. *Nat. Rev. Phys.* **6**, 342–344.

KRENN, M., *et al.* 2022 On scientific understanding with artificial intelligence. *Nat. Rev. Phys.* **4**, 761–769.

LAKSHMINARAYANAN, B., PRITZEL, A. & BLUNDELL, C. 2017 Simple and scalable predictive uncertainty esimation using deep ensembles. In *Advanced in Neural Information Processing Systems*.

LAM, R., SANCHEZ-GONZALEZ, A., WILLSON, M., WIRNSBERGER, P., FORTUNATO, M., ALET, F., RAVURI, S., EWALDS, T., EATON-ROSEN, Z. & BATTAGLIA, P. 2023 Learning skillful medium-range global weather forecasting. *Science* **382** (6677), 1416–1421.

LAMBERT, D. 1997 L'efficacité des mathématiques est-elle déraisonnable. *Bull. Acad. R. Belgique* **8-1-6**, 161–178.

LAROCHELLE, H., ERHAN, D., COURVILLE, A.C., BERGSTRA, J. & BENGIO, Y. 2007 An empirical evaluation of deep architectures on problems with many factors of variation. In *ICML*, pp. 473–480.

LAV, C., BANKO, A.J., WASCHKOWSKI, F., ZHAO, Y., ELKINS, C.J., EATON, J.K. & SANDBERG, R.D. 2023 A coupled framework for symbolic turbulence models from deep-learning. *Intl J. Heat Fluid Flow* **101**, 109140.

LEI, W.-M. & LI, H.-B. 2025 U-WNO: U-Net enhanced wavelet neural operator for solving parametric partial differential equations. *Comput. Maths Applics* **194**, 272–287.

LEMOS, P., JEFFREY, N., CRANMER, M., HO, S. & BATTAGLIA, P. 2023 Rediscovering orbital mechanics with machine learning. *Mach. Learn.: Sci. Technol.* **4**, 045002.

LEVINS, R. 1966 The strategy of model building in population biology. *Am. Sci.* **54** (4), 421–431.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

LI, Q., HAN, Z. & WU, X. 2018 Deeper insights into graph convolutional networks for semi-supervised learning. *AAAI Conf. Artif. Intell.* **32**, 3538–3545.
LI, T., BUZZICOTTI, M., BIFERALE, L. & BONACCORSO, F. 2023 Generative adversarial networks to infer velocity components in rotating turbulent flows. *Eur. Phys. J. E* **46**, 31.
LI, Z., KOVACHKI, N., AZIZZADENESHELI, K., LIU, B., BHATTACHARYA, K., STUART, A. & ANANDKUMAR, A. 2021 Fourier neural operator for parametric partial differential equations. In *Proceedings of ICLR*.
LINDEBERG, T. 2013 Invariance of visual operations at the level of receptive fields. *PLoS ONE* **8** (7), e66990.
LINO, M., FOTIADIS, S., BHARATH, A.A. & CANTWELL, C.D. 2023 Current and emerging deep-learning methods for the simulation of fluid dynamics. *Proc. R. Soc. Lond. A* **479**, 2275.
LIPTON, Z.C. 2016 The mythos of model interpretability. In *ICML Workshop on Human Interpretability in Machine Learning*.
LIU, D. & WANG, Y. 2019 Multi-fidelity physics-constrained neural network and its application in materials modeling. In *Proceedings of the ASME 2019 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 2A. V02AT03A007.
LIU, H.-T.D., WILLIAMS, F., JACOBSON, A., FIDLER, S. & LITANY, O. 2022*a* Learning smooth neural functions via Lipschitz regularization. In *SIGGRAPH*.
LIU, X., YOO, C., XING, F., OH, H., FAKHRI, G.EL, KANG, J.-W. & WOO, J. 2022*b* Deep unsupervised domain adaptation: a review of recent advanced and perspectives. *APSIPA Transactions on Signal and Informaiton Processings*, vol. 1, pp. 1–48.
LIU, Y., SUN, J., HE, X., PINNEY, G., ZHANG, Z. & SCHAEFFER, H. 2024 PROSE-FD: a multimodal PDE foundation model for learning multiple operators for forecasting fluid dynamics. arXiv 2409.09811v1.
LODATO, G. 2019 Characteristic modal shock detection for discontinuous finite element methods. *Comput. Fluids* **179**, 309–333.
LU, L., JIN, P. & KARNIADAKIS, G.E. 2019 Deeponet: learning nonlinear operators for identifying differential equations based on the universal approximation theorem of operators. arXiv 1910.03193.
LU, L., MENG, X., MAO, Z. & KARNIADAKIS, G.E. 2021 DeepXDE: a deep learning library for solving differential equations. *SIAM Rev.* **63** (1), 208–228.
LUIGI, L.DE, CARDACE, A., SPEZIALETTI, R., RAMIREZ, P.Z., SALTI, S. & STEFANO, L.DI 2023 Deep learning on implicit neural representations of shapes. In *ICLR*.
LUSCH, B., KUTZ, J.N. & BRUNTON, L. 2018 Deep learning for universal linear embeddings of nonlinear dynamics. *Nat. Commun.* **9**, 4950.
LÜTJENS, B., CRAWFORD, C.H., WATSON, C., HILL, C. & NEWMAN, D. 2022 Multiscale neural operator: learning fast and grid-independent PDE solvers. In *AI4Science Workshop from ICML*, vol. 39.
LYU, Y., ZHAO, X., GONG, Z., KANG, X. & YAO, W. 2023 Multi-fidelity prediction of fluid flow based on transfer learning using Fourier neural operator. *Phys. Fluids* **35** (7), 077118.
MACKAY, D.J. 1995 Bayesian neural networks and density networks. *Nucl. Instrum. Meth. Phys. Res. A Accelerators Spectrometers Detectors Associated Equipment* **354** (1), 73–80.
MADANI, A., *et al.* 2023 Large language models generate functional protein sequences across diverse families. *Nat. Biotechnol.* **41**, 1099–1106.
MAGRI, L., SEE, Y.-C., TAMMISOLA, O., IHME, M. & JUNIPER, M. 2016 Multiple-scale thermoacoustic stability analysis of a coaxial jet combustor. In *Proceedings of the Combustion Institute* (06.009).
MALLAT, S. 2009 *A Wavelet Tour of Signal Processing*. The sparse way.
MARIÉ, S., RICOT, D. & SAGAUT, P. 2009 Comparison between lattice Boltzmann method and Navier–Stokes high order schemes for computational aeroacoustics. *J. Comput. Phys.* **228** (4), 1056–1070.
MARKELLE, K., LONGJOHN, R. & NOTTINGHAM, K. 2014 The UCI machine learning repository.
MARONNE, S. & PATRAS, F. 2022 L'épistémologie mathématique de Gaston Bachelard. *Bachelard Stud.* **1–2**, 51–68.
MATHIEU, M., COUPRIE, C. & LECUN, Y. 2016 Deep multi-scale video prediction beyond mean square error. In *ICLR*.
MATSUMOTO, D. & NAKAI, T. 2023 Syntactic theory of mathematical expressions. *Cognitive Psychol.* **146**, 101606.
MAXWELL, J.C. 1861 On physical lines of force. *Phil. Mag. J. Sci.* **21**, 488.
MCCARTHY, J. 1981 Epistemological problems of artificial intelligence. In *Readings in Artificial Intelligence*, pp. 459–465.
MCCARTHY, J., MINSKY, M.L., ROCHESTER, N. & SHANNON, C.E. 1956 A proposal for the Dartmouth summer research project on artificial intelligence. Availavle at: https://raysolomonoff.com/dartmouth/20boxa/dart564props.pdf.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

McCulloch, W.S. & Pitts, W. 1943 A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **5**, 115–133.
Meneveau, C. & Lund, T. 1997 The dynamic Smagorinsky model and scale-dependent coefficients in the viscous range of turbulence. *Phys. Fluids* **9** (12), 3932–3934.
Meneveau, C. & Sreenivasan, K.R. 1987 The multifractal spectrum of the dissipation field in turbulent flows. *Nucl. Phys. B Proc. Suppl.* **2**, 49.
Menon, K. & Mittal, R. 2021 Significance of the strain-dominated region around a vortex on induced aerodynamic loads. *J. Fluid Mech.* **918**, R3.
Michel, T., Cvjetko, M., Hamon, G., Oudeyer, P.-Y. & Moulin-Frier, C. 2025 Exploring flow-lenia universes with a curiosity-driven AI scientist: discovering diverse ecosystem dynamics. arXiv 2505.15998v2.
Minisci, E. & Vasile, M. 2013 Robust design of a reentry unmanned space vehicle by multifidelity evolution control. *AIAA J.* **51** (6), 1284–1295.
Mitchell, M. 1996 *An Introduction to Genetic Algorithms*. Cambridge MIT Press.
Moser, P., Fenz, W., Thumfart, S., Ganitzer, I. & Giretzlehner, M. 2023 Modeling of 3D blood flows with physics-informed neural networks: comparison of network architectures. *Fluids* **8** (2), 46.
Müller, B. 2020 Hydrodynamics of core-collapse supernovae and their progenitors. *Living Rev. Comput. Astrophys.* **6**, 3.
Müller, T., Evans, A., Schied, C. & Keller, A. 2022 Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.* **41** (4), 1–15.
Nabian, M.A. 2014 X-meshgraphnet: scalable multi-scale graph neural networks for physics simulation. arXiv 2411.17164v1.
Nabian, M.A. & Meidani, H. 2020 Physics-driven regularization of deep neural networks for enhanced engineering design and analysis. *J. Comput. Inform. Sci. Engng* **20** (1), 011006.
Najar, M.Al, Almar, R., Bergsma, E.W.J., Delvit, J.-M. & Wilson, D.G. 2023 Improving a shoreline forecasting model with symbolic regression. In *Tackling Climate change with Machine Learning, ICLR*, p. hal–04281530.
Nicoud, F. 2004 High-order conservative schemes for low Mach number flows. In *Tenth SIAM International Conference on Numerical Combustion*.
Nicoud, F., Baggett, J., Moin, P. & Cabot, W. 2001 LES wall-modeling based on optimal control theory. *Phys. Fluids* **13** (10), 1629–1632.
Nicoud, F. & Ducros, F. 1999 Subgrid-scale stress modelling based on the square of the velocity gradient. *Flow Turbul. Combust.* **62** (3), 183–200.
Noether, E. 1918 Invariante variationsprobleme. In *Nachrichten von der Gesellschaft der Wissenschaften zu Go¨ttingen. Mathematisch-Physikalische Klasse*, pp. 235–257.
O'Connor, J., Day, P., Mandal, P. & Revell, A. 2016 Computational fluid dynamics in the microcirculation and microfluidics: what role can the lattice Boltzmann method play? *Integr. Biol.* **8** (5), 589–602.
Ohana, R., *et al.* 2024 The well: a large-scale collection of diverse physics simulations for machine learning. In *NeurIPS*, vol. 38.
Önder, A. & Liu, P.L.-F. 2023 Deep learning of interfacial curvature: a symmetry-preserving approach for the volume of fluid method. *J. Comput. Phys.* **485**, 112110.
Otto, S.E., Zolman, N., Kutz, J.N. & Brunton, S.L. 2023 A unifified framework to enforce, discover and promote symmetry in machine learning. arXiv 2311.00212.
Ovenden, N.C. 2005 A uniformly valid multiple scales solution for cut-on cut-off transition of sound in flow ducts. *J. Sound Vibr.* **286** (1–2), 403–416.
Pan, S.J. & Yang, Q. 2010 A survey on transfer learning. *IEEE Trans. Knowledge Data Engng* **22** (10), 1345–1359.
Patterson, R.E. & Eggleston, R.G. 2017 Intuitive cognition. *J. Cognitive Engng Decision Making* **11** (1), 5–22.
Peherstorfer, B., Willcox, K. & Gunzburger, M. 2018 Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Rev.* **60** (3), 550–591.
Penrose, R. 1990 *The Emperor's New Mind: Concerning Computers, Minds, and The Laws of Physics*. Oxford University Press.
Pinto, W.G., Alguacil, A. & Bauerheim, M. 2022 On the reproducibility of fully convolutional neural networks for modeling time–space-evolving physical systems. *Data-Centric Engng* **3**, e19.
Popper, K.R. 1962 *Conjectures and Refutations: The Growth of Scientific Knowledge*. Routledge.
Pratt, W.K. 1972 Generalized wiener filtering computation techniques. *IEEE Trans. Comput.* **100** (7), 636–641.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

QIU, J., HUANG, J., ZHANG, X., LIN, Z., PAN, M., LIU, Z. & MIAO, F. 2024 Pi-fusion: physics-informed diffusion model for learning fluid dynamics. arXiv2024 2406.03711v1.

QUEDEVILLE, V. 2020 Mathematical analysis, modelling and simulation of microbial population dynamics. PhD thesis, Institut National Polytechnique de Toulouse, France.

RABAULT, J., KUCHTA, M., JENSEN, A., RÉGLADE, U. & CERARDI, N. 2019 Artificial neural networks trained through deep reinforcement learning discover control strategies for active flow control. *J. Fluid Mech.* **865**, 281–302.

RAISSI, M., PERDIKARIS, P. & KARNIADAKIS, G.E. 2019 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.

RAZAFINDRALANDY, D., HAMDOUNI, A. & OBERLACK, M. 2007 New turbulence models preserving symmetries. In *5th International Symposium on Turbulence and Shear Flow Phenomen*.

REISSMANN, M., HASSLBERGER, J., SANDBERG, R.D. & KLEIN, M. 2021 Application of gene expression programming to a-posteriori LES modeling of a Taylor Green vortex. *J. Comput. Phys.* **424**, 109859.

ROBIC, S. & JUNGCK, J.R. 2011 Unraveling the tangled complexity of DNA: combining mathematical modeling and experimental biology to understand replication, recombination and repair. *Math. Model. Natural Phenom.* **6** (6), 108–135.

RONNEBERGER, O., FISCHER, P. & BROX, T. 2015 U-net: convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015*, pp. 234–241.

ROSENBLATT, F. 1958 The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **65** (6), 386–408.

ROY, P.T., SEGUI, L.M., JOUHAUD, J.C. & GICQUEL, L. 2018 Resampling strategies to improve surrogate model-based uncertainty quantification: application to LES of LS89. *Intl J. Numer. Meth. Fluids* **87** (12), 607–627.

RUDY, S.H., BRUNTON, S.L., PROCTOR, J.L. & KUTZ, J.N. 2017 Data-driven discovery of partial differential equations. *Sci. Adv.* **3**, e1602614.

RUSCH, T.K., BRONSTEIN, M.M. & MISHRA, S. 2023 A survey on oversmoothing in graph neural networks. arXiv 2303.10993v1.

RUSSAKOVSKY, O., *et al.* 2015 Imagenet large scale visual recognition challenge. *Intl J. Comput. Vis.* **115**, 211–252.

RYWIK, M., ZIMMERMANN, A., EDER, A.J., SCOLETTA, E. & POLIFKE, W. 2023 Spatially resolved modeling of the nonlinear dynamics of a laminar premixed flame with a multilayer perceptron - convolution autoencoder network. In *Proceedings of ASME Turbo Expo*, p. 102543.

SA, A.A.R.D., CARVALHO, J.D. & NAVES, E.L.M. 2023 Reflections on epistemological aspects of artificial intelligence during the COVID-19 pandemic. *AI Soc.* **38**, 1601–1608.

SAETTA, E., TOGNACCINI, R. & IACCARINO, G. 2024 Uncertainty quantification in autoencoders predictions: applications in aerodynamics. *J. Comput. Phys.* **506**, 112951.

SAYED, N.A., HAMDOUNI, A., LIBERGE, E. & RAZAFINDRALANDY, D. 2010 The symmetry group of the non-isothermal Navier–Stokes equations and turbulence modelling. *Symmetry* **2** (2), 848–867.

SCHMID, P.J. 2010 Dynamic mode decomposition of numerical and experimental data. *J. Fluid Mech.* **656**, 5–28.

SEITZLER, M., TAVAKOLI, A., ANTIC, D. & MARTIUS, G. 2022 On the pitfalls of heteroscedastic uncertainty estimation with probabilistic neural networks. In *ICLR*.

SERRANO, L., WANG, T.X., NAOUR, E.LE, VITTAUT, J.-N. & GALLINARI, P. 2024 Aroma: preserving spatial structure for latent PDE modeling with local neural fields. In *NeurIPS*, vol. 38.

SERRURIER, M., MAMALET, F., GONZALEZ-SANZ, A., BOISSIN, T., LOUBES, J.-M. & BARRIO, E.D. 2021 Achieving robustness in classification using optimal transport with hinge regularization. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 505–514.

SHU, D., LI, Z. & FARIMANI, A.B. 2023 A physics-informed diffusion model for high-fidelity flow field reconstruction. *J. Comput. Phys.* **478**, 111972.

SIDDANI, B., BALACHANDAR, S. & FANG, R. 2021 Rotational and reflectional equivariant convolutional neural network for data-limited applications: multiphase flow demonstration. *Phys. Fluids* **33**, 103323.

SIERRA-AUSIN, J., FABRE, D. & KNOBLOCH, E. 2024 Wake dynamics in buoyancy-driven flows: steady-state-Hopf-mode interaction with O(2) symmetry revisited. *Phys. Rev. E* **109** (1), 014216.

SILVA, B.M.D., HIGDON, D.M., BRUNTON, S.L. & KUTZ, J.N. 2020 Discovery of physics from data: universal laws and discrepancies. *Frontiers Artif. Intell.* **3**, 25.

SIMON, H.A. 1962 The architecture of complexity. *Proc. Am. Phil. Soc.* **106** (6), 467–482.

SIMON, H.A. 1969 *The Sciences of the Artificial*. MIT Press.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

SIMONDON, G. 1970 *Naissance de la technologie*. Sur la Technique.

SITZMANN, V., MARTEL, J.N.P., BERGMAN, A.W., LINDELL, D.B. & WETZSTEIN, G. 2020 Implicit neural representations with periodic activation functions. In *Conference on Neural Information Processing Systems*, vol. 34.

SMAGORINSKY, J. 1963 General circulation experiments with the primitive equations: 1. The basic experiment. *Mon. Weath. Rev.* **91**, 99–164.

SONG, J., MENG, C. & ERMON, S. 2020 Denoising diffusion implicit models. arXiv 2010.02502.

SUBRAMANIAN, S., HARRINGTON, P., KEUTZER, K., BHIMJI, W., MOROZOV, D., MAHONEY, M.W. & GHOLAMI, A. 2023 Towards foundation models for scientific machine learning: characterizing scaling and transfer behavior. In *NeurIPS*, vol. 37.

TANCIK, M., SRINIVASAN, P.P., MILDENHALL, B., FRIDOVICH-KEIL, S., RAGHAVAN, N., SINGHAL, U., RAMAMOORTHI, R., BARRON, J.T. & NG, R. 2020 Fourier features let networks learn high frequency functions in low dimensional domains. In *Conference on Neural Information Processing Systems*, vol. 34.

TARTAGLIONE, E., LATHUILIÈRE, S., FIANDROTTI, A., CAGNAZZO, M. & GRANGETTO, M. 2021 HEMP: high-order entropy minimization for neural network compression. *Neurocomputing* **461**, 244–253.

TEKKESIN, A.I. 2019 Artificial intelligence in healthcase: past, present and future. *Anatolian J. Cardiology* **22** (2), 8–9.

TENACHI, W., IBATA, R. & DIAKOGIANNIS, F.I. 2023 Deep symbolic regression for physics guided by units constraints: toward the automated discovery of physical laws. *Astrophys. J.* **959** (2), 99.

THELEN, A.S., BRYSON, D.E., STANFORD, B.K. & BERAN, P.S. 2022 Multi-fidelity gradient-based optimization for high-dimensional aeroelastic configurations. *Algorithms* **15** (4), 131.

THOM, R., TSATSANIS, S.P. & LISKER, R. 2016 *To Predict IS NOT To Explain: Conversations On Mathematics, Science, Catastrophe Theory, Semiophysics, Morphogenesis and Natural Philosophy*. Thombooks Press.

TIERZ, A., IPARRAGUIRRE, M.M., ALFARO, I., GONZALEZ, D., CHINESTA, F. & CUETO, E. 2024 On the feasibility of foundational models for the simulation of physical phenomena. arXiv 2410.14645v1.

TOMPSON, J., SCHLACHTER, K., SPRECHMANN, P. & PERLIN, K. 2017 Accelerating Eulerian fluid simulation with convolutional networks. In *Proceedings of the 34th ICML*.

TOUBOUL, P. 2022 Microscope mission: final results of the test of the equivalence principle. *Phys. Rev. Lett.* **129**, 121102.

TOUVRON, H., *et al.* 2023 LLaMA: open and efficient foundation language models. arXiv 2302.13971v1.

TRASK, N., HUANG, A. & HU, X. 2022 Enforcing exact physics in scientific machine learning: a data-driven exterior calculus on graphs. *J. Comput. Phys.* **456**, 110969.

TRIPURA, T. & CHAKRABORTY, S. 2023 Wavelet neural operator for solving parametric partial differential equations in computational mechanics problems. *Comput. Meth. Appl. Mech. Engng* **404**, 115783.

UM, K., BRAND, R., FEI, Y., HOLL, P. & THUEREY, N. 2020 Solver-in-the-loop: learning from differentiable physics to interact with iterative PDE-solvers. In *NeurIPS*, vol. 34.

USANDIVARAS, J.F.Z., BAUERHEIM, M., CUENOT, B. & URBANO, A. 2024*a* Data-driven multi-fidelity surrogate models for rocket engines injector design. *Data-Centric Engng* (under review).

USANDIVARAS, J.F.Z., BAUERHEIM, M., CUENOT, B. & URBANO, A. 2024*b* Deep convolutional autoencoders for the time–space reconstruction of liquid rocket engine flames. *Proc. Combust. Inst.* **40** (1–4), 105382.

VANDAMME, F. 1976 L'épistémologie et la sémantique. *Revue Internationale de Philosophie* **117** (118), 435–449.

VASWANY, A., SHAZEER, N., PARMAR, N., USZKOREIT, J., JONES, L., GOMEZ, A.N. & KAISER, L. 2017 Attention is all you need. In *NIPS*, vol. 31.

VELTHOVEN, M. & MARCUS, E. 2024 Problems in AI, their roots in philosophy, and implications for science and society. arXiv 2407.15671.

VERMA, S., NOVATI, G. & KOUMOUTSAKOS, P. 2018 Efficient collective swimming by harnessing vortices through deep reinforcement learning. *Proc. Natl Acad. Sci. USA* **115** (23), 5849–5854.

WANG, R., WALTERS, R. & YU, R. 2021 Incorporating symmetry into deep dynamics models for improved generalization. In *ICLR*.

WANG, Y., LI, Z., YUAN, Z., PENG, W., LIU, T. & WANG, J. 2024 Prediction of turbulent channel flow using Fourier neural operator-based machine-learning strategy. *Phys. Rev. Fluids* **9** (8), 084604.

WEI, Z., DUFOUR, E.R., PELLETIER, C., FUA, P. & BAUERHEIM, M. 2024 DiffAirfoil: an efficient novel airfoil sampler based on latent space diffusion model for aerodynamic shape optimization. In *AIAA Aviation Forum*, p. 3755.

WEILER, M. & CESA, G. 2019 General E(2) - equivariant steerable CNNs. In *33rd Conference on Neural Information Processing Systems*.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press

WEILER, M., FORRÉ, P., VERLINDE, E. & WELLING, M. 2023 *Equivariant and Coordinate Independent Convolutional Networks*. University of Amsterdam.

WEILER, M., GEIGER, M., WELLING, M., BOOMSMA, W. & COHEN, T.S. 2018 3D steerable CNNs: learning rotationally equivariant features in volumetric data. *Adv. Neural Inform. Proc. Syst.* **31**.

WIGNER, E. 1960 The unreasonable effectiveness of mathematics in natural sciences. *Commun. Pure Appl. Maths* **13** (1), 1–14.

WU, C., ZHU, M., TAN, Q., KARTHA, Y. & LU, L. 2023 A comprehensive study of non-adaptive and residual-based adaptive sampling for physics-informed neural networks. *Comput. Meth. Appl. Mech. Engng* **403** (A), 115671.

WU, Z., JAIN, P., WRIGHT, M.A., MIRHOSEINI, A., GONZALEZ, J.E. & STOICA, I. 2021 Representing long-range context for graph neural networks with global attention. In *Conference on Neural Information Processing Systems*, vol. 35.

XIA, S., XIONG, Z., LUO, Y., XU, X. & ZHANG, G. 2015 Effectiveness of the Euclidean distance in high dimensional spaces. *Optik* **126** (24), 5614–5619.

XIE, Y., FRANZ, E., CHU, M. & THUEREY, N. 2018 Tempogan: a temporally coherent, volumetric gan for super-resolution fluid flow. *ACM Trans. Graph.* **37** (4), 1–15.

XING, V. & LAPEYRE, C.J. 2023 *Deep Convolutional Neural Networks for Subgrid-Scale Flame Wrinkling*. Springer International Publishing.

YANG, J., ZHOU, K., LI, Y. & LIU, Z. 2024 Generalized out-of-distribution detection: a survey. *Intl J. Comput. Vis.* **132**, 5635–5662.

YIN, Y., KIRCHMEYER, M., FRANCESCHI, J.-Y., RAKOTOMAMONJY, A. & GALLINARY, P. 2023 Continuous PDE dynamics forecasting with implicit neural representations. In *ICLR*.

YOSHIDA, H. 1990 Construction of higher order symplectic integrators. *Phys. Lett. A* **150** (5–7), 262–268.

YUAN, L., *et al.* 2021 Florence: a new foundation model for computer vision. arXiv 2111.11432v1.

ZHANG, L. & SCHAEFFER, H. 2019 On the convergence of the SINDy algorithm. *SIAM Multiscale Model. Simul.* **17** (3), 948–972.

ZHANG, M., KIM, S., LU, P.Y. & SOLJACIC, M. 2023 Deep learning and symbolic regression for discovering parametric equations. *IEEE Trans. on Neural Networks Learn. Syst.* **35** (11), 16775–16787.

ZHENG, P., ASKHAM, T., BRUNTON, S.L., KUTZ, J.N. & ARAVKING, A.Y. 2018 A unified framework for sparse relaxed regularized regression: Sr3. *IEEE Access* **7**, 1404–1423.

ZHOU, J., CUI, G., HU, S., ZHANG, Z., YANG, C., LIU, Z., WANG, L., LI, C. & SUN, M. 2020 Graph neural networks: a review of methods and applications. *AI Open* **1**, 57–81.

https://doi.org/10.1017/jfm.2026.11256 Published online by Cambridge University Press