

POSITION PAPER

Machine learning and partial differential equations: benchmark, simplify, and discover

Petros Koumoutsakos 

Computational Science and Engineering Laboratory, Harvard University, Cambridge, MA 02138, USA
Email: petros@seas.harvard.edu

Received: 29 September 2024; **Revised:** 24 January 2025; **Accepted:** 19 February 2025

Keywords: generative AI; machine learning; PDEs; scientific computing

Abstract

Simulations of critical phenomena, such as wildfires, epidemics, and ocean dynamics, are indispensable tools for decision-making. Many of these simulations are based on models expressed as Partial Differential Equations (PDEs). PDEs are invaluable inductive inference engines, as their solutions generalize beyond the particular problems they describe. Methods and insights acquired by solving the Navier–Stokes equations for turbulence can be very useful in tackling the Black-Scholes equations in finance. Advances in numerical methods, algorithms, software, and hardware over the last 60 years have enabled simulation frontiers that were unimaginable a couple of decades ago. However, there are increasing concerns that such advances are not sustainable. The energy demands of computers are soaring, while the availability of vast amounts of data and Machine Learning (ML) techniques are challenging classical methods of inference and even the need of PDE based forecasting of complex systems. I believe that the relationship between ML and PDEs needs to be reset. PDEs are not the only answer to modeling and ML is not necessarily a replacement, but a potent companion of human thinking. Algorithmic alloys of scientific computing and ML present a disruptive potential for the reliable and robust forecasting of complex systems. In order to achieve these advances, we argue for a rigorous assessment of their relative merits and drawbacks and the adoption of probabilistic thinking for developing complementary concepts between ML and scientific computing. The convergence of AI and scientific computing opens new horizons for scientific discovery and effective decision-making.

Impact statement

Modeling and solving complex systems based on Partial Differential Equations (PDE) and machine learning (ML) can unlock a new era of robust and reliable forecasting for complex systems such as wildfires, epidemics, and climate change. This interdisciplinary approach promises to revolutionize scientific discovery and empower effective decision-making in critical areas.

1. Introduction

“As all of you are, I am concerned about the world in which we are going to live tomorrow, a world in which a new machine, the digital computer, may be of even greater importance than the atomic bomb.”

These are the words of screen actor David Wayne in a chat with Professor Jerome Wiesner (MIT) on developments in Artificial Intelligence (AI) that aired on the US television network CBS in 1961 (it can be found at this link <https://techtv.mit.edu/videos/10268-the-thinking-machine-1961---mit-centennial->

film). This prediction proved prescient, as 30 years later, following the ban on system-level nuclear testing, computer simulations were proposed as a viable alternative to physical bomb detonations. Simulations have indeed become a key tool in the U.S. nuclear stockpile maintenance program. In the 30 years since the ban, researchers around the world have raced to make such simulations a reality, leading to immense progress in simulation capabilities of complex systems across scientific domains. A key enabler has been the exponential growth in transistor density on microchips, roughly doubling every two years, which has significantly increased computational power while reducing relative costs. This growth, described as Moore's law, forged new connections between disciplines and laid the foundations of computational science, which is now widely recognized today as the third pillar of scientific inquiry, along with theory and experiments. Today, predictive simulations with quantified uncertainties of critical phenomena described by PDEs, such as climate and epidemics, are becoming a key element in human decision-making and engineering design.

However, many experts believe that Moore's law is nearing its end, or at least slowing down, due to physical and economic constraints (Shalf, 2020). There are major concerns about the energy demands of computers and the veracity and reliability of computer simulations. At the same time, new hope for modeling and predicting complex systems is emerging, driven by recent advances in Machine Learning (ML) and AI. The idea of using artificial intelligence (AI) for modeling complex systems has a long history as a part of the semi-forgotten field of cybernetics (Wiener, 1950; Tsetlin, 1961; Ivanenko and Lapa, 1967). However, what is novel in our times is the trillion-fold increase in hardware capabilities and the dizzying pace of acquiring, transmitting, and processing massive amounts of data. Despite these unprecedented capabilities, there is broad consensus that resolving the full range of spatiotemporal scales in complex systems such as turbulence, ocean circulation, and cloud dynamics will remain out of reach for the foreseeable future (Palmer, 2015). An alternative is reduced-order models (ROMs) that may bypass the need for expensive simulations. ML and, more broadly, AI methodologies can be invaluable tools for constructing data-driven ROMs that may also serve as companions to expensive simulations. Indeed, as we are experiencing unprecedented capabilities of computers, new forms of computational thinking emerge that were unimaginable only a couple of decades ago. One example is Digital Twins, which have become the epitome and frontier of our computing capabilities. DTs play an essential role in the intricate tasks of designing, constructing, operating, and maintaining complex systems across disciplines ranging from medicine to aircraft design (Glaessgen and Stargel, 2012; Venkatesh et al., 2022; Ferrari and Willcox, 2024). DTs are central to decision-making frameworks, which often rely on designs tailored to specific cases and may lack the flexibility needed to handle situations with limited data or unclear system dynamics. We believe that there are open horizons in the embedding of DTs within decision-making frameworks. We envision probabilistic DTs that go beyond PDEs but employ multifidelity models and machine learning, while adhering to physical constraints and their predictive uncertainties that are continuously updated with available data over different time horizons to support diverse decision-making scenarios. At the same time, new concepts such as foundation models and generative AI open up new vistas for DTs and beyond that need to be explored (Bommasani et al., 2021; Jacobsen et al., 2023; Bodnar et al., 2024; Li et al., 2024; Price et al., 2024; Gao et al., 2024a). The perspective of researchers familiar with PDEs can be invaluable in elucidating the capabilities and limitations of foundational models.

In this article, I share a personal perspective shaped by 25 years of experience (Müller et al., 1999; Brunton et al., 2020) in solving complex PDE problems using scientific computing and machine learning, discussing successes, failures, and the balance between hype and hope at the ML-PDE interface. With all due respect to the numerous colleagues who have influenced my work, this perspective is primarily reflected in the publications of our group in Neural Networks (NN) for evolving systems described by PDEs, Reinforcement Learning for closures of ROMs, and recent work on generative AI for capturing the evolution of complex physical systems. I advocate for the creation of algorithmic alloys that combine scientific computing and AI to tackle some of the most challenging problems of our times. I believe that a particularly attractive feature of ML is its inherently statistical nature that allows for a systematic quantification of uncertainty. By forming algorithmic alloys with scientific computing, we obtain access

to robust ways of computing the uncertainty of the predictions obtained from solving PDEs. I believe that AI not only enhances automation by maximizing machine capabilities but also unlocks new paths for human thought and scientific discovery.

2. Benchmarking is critical when comparing scientific computing and machine learning

Computing can be understood as a process of dimensionality reduction, where infinite or high-dimensional fields are translated into reduced-order, query-driven information. This perspective aligns with computations ranging from forecasting hurricane trajectories in scientific computing to recognizing cats in image collections through machine learning. In order to juxtapose computations made in these fields, I use an example that I first saw in a talk by Christopher Bishop (Microsoft). One of the most well-known algorithms for dimensionality reduction is the linear principal component (PCA) analysis. The N dimensional data vector is used to form a covariance matrix, and extracting its $M \ll N$ eigenvectors that correspond to its largest eigenvalues provide a reduced-order basis for representing the data. However, venturing beyond the linear representation offered by PCA is not a straightforward task. The task of developing a non-linear PCA is much easier if one adopts a neural network (NN) architecture. A single-layer NN with linear activation functions that learns the identity mapping is a linear autoencoder equivalent to linear PCA (Baldi and Hornik, 1989). Unlike the eigenvalue problem, it is possible to render the NN nonlinear by simply changing the node activation functions. Adding more layers allows for more flexibility and more powerful nonlinear dimensionality reduction algorithms. This was the architecture we adopted in order to compare dimensionality reduction using NN and Proper Orthogonal Decomposition (POD- the name with which linear PCA is known in the fluid mechanics community) (Berkooz et al., 1993). To the best of my knowledge, this was the first application of deep neural networks (DNN) in fluid mechanics and turbulent flows (Milano and Koumoutsakos, 2002). In that same paper, NNs complemented an analytical expression for reconstructing the flow above the wall using only wall-based information suggests the hybridization of numerical/analytical approaches with and ML. One of the reviewers of that same paper raised the issue of interpretability in using DNN encodings at the time. Indeed, interpreting the behavior of DNNs is not as straightforward as for a linear eigenvalue problem. But again, interpretability is a human trait, and this view may only apply to researchers trained in classical numerical analysis concepts. New generations of scientists may be more comfortable with DNNs and their ablation studies than with linear stability analysis. Furthermore, training for DNNs may become stuck in local minima of the objective function. Training a DNN 25 years ago was 7–8 orders of magnitude slower than it is today due to the speed of computers (2 minutes versus a year for the same DNN based on the speed of the top supercomputer of the Top500 list). I believe that several interesting ideas in ML for PDEs today owe their inception to these capabilities.

In recent years, ML algorithms have exploited the ample availability of data and powerful computing architectures (Kurth et al., 2018) to provide new solutions to problems across all scientific fields (Baldi et al., 2014; Carleo et al., 2019; Kochkov et al., 2024). The advent of large language models adds yet another dimension to the capabilities of human discovery (Buehler, 2024; Meng et al., 2024). This excitement is unavoidably accompanied by a certain degree of hype on the potential of AI and Machine Learning (ML) as a replacement for scientific computing. Indeed, ML has shown potential for bypassing classical scientific computing and the forecasting of systems based on PDEs by taking the alternate route of data-driven modeling. Successes in climate and weather forecasting are a testament to this capability (Lam et al., 2023; Bodnar et al., 2024; Price et al., 2024). At the same time, there have been limited advances in solving challenging PDEs using ML methods. Efforts to adopt (or plug in) ML techniques in scientific computing frameworks (such as using NNs for Galerkin type of numerical methods) are attractive due to the relative simplicity of using a cost function and the availability of automatic differentiation tools. Furthermore, it is more attractive these days to use NNs in the title of a paper instead of Galerkin. However, this fusion may not be the final answer in forecasting complex systems. Several popular methods (Bar-Sinai et al., 2019; Li et al., 2020; Karniadakis et al., 2021) have solved a plethora of low-dimensional, simplified problems, but extensions to complex multiscale systems, such as

flows at high Reynolds numbers, remain an open question. Hence, I wonder if for future research, the right direction is to try to replace, for example, complex flux limiters and shock-capturing schemes with deep NNs to solve supersonic flows. There are limitations of ML learning methods in forecasting, for example, the evolution of the complete (near body and wake) vorticity field in flows past a circular cylinder or vortex merging at Reynolds numbers above 5000, even in 2D. This fact does not discount the current developments but suggests that more potent computer hardware is an enabler for innovation and new modes of inquiry in human thinking. Today, ML offers new impulses, hope (mixed with some hype) for solving PDEs across disciplines. However, in order to achieve progress in understanding and forecasting complex systems, we need to specify rigorous metrics that quantify the veracity and robustness of ML solutions for PDEs (McGreivy and Hakim, 2024). There is an ongoing debate about whether the *physics-informed* approaches are superior to the *physics-consistent* approaches. The latter have been ingrained for decades in the development of numerical methods, but indeed, they can be dauntingly complex. The former promises simplicity along with abundant software. Which direction should we choose? I believe that to decide on a path forward, we need a critical evaluation of these approaches. The concepts of validation and verification and uncertainty quantification that have become the cornerstones of scientific computing are making inroads in machine learning (Gal et al., 2022). It would be very useful if the ROMs developed through AI or scientific computing techniques were accompanied with guarantees and error bounds of their predictions. Although consistency with first-principles models is not at the core of data-driven approaches, the use of synthetic data from solving PDEs can guide the development of guarantees for ML methods.

Perhaps, instead of trying to answer the question “can AI replace numerical methods?” we need to reflect more on whether this is the right question to ask. Probabilities are a powerful framework to base this discussion (Jaynes, 2003). In a departure from deterministic solutions of PDEs, the outcomes of ML are statistical objects that we can use to quantify the element of surprise in our assumptions. Bayesian model selection suggests mixture models (possibly based on PDEs) with quantified probabilities as weights. In case a single model needs to be selected, Bayesian inference offers relevant methods (Zavadlav et al., 2019). In turn, the fusion of ROMs and expensive simulations may offer the best of both worlds. Generative models can sample data from high-dimensional target distributions and can be conditioned on specific quantities or latent structures (Gao et al., 2024a). This path provides exciting opportunities for a novel way of handling Uncertainty Quantification (UQ) in systems described by PDEs.

3. Simplify interfaces: patterns across scales in space and time

The growth of ML has led to a “gold rush” across all scientific fields. As the number of publications increases at an exponential pace, it is sometimes difficult to identify novelty, in particular, in advances occurring at the important interfaces between disciplines. I believe that the simplification of questions, algorithms, and results can facilitate the identification of common elements at the interface of AI and scientific computing. The breakthrough in ML in pattern recognition came with the publication of the ImageNet classification achieved by a convolutional neural network (Krizhevsky et al., 2012). Today, it is well-known that ML has an advantage in pattern recognition across a broad range of feature spaces (Theodoridis and Koutroumbas, 2006). Patterns are also present in dynamical systems, not only in space and time but also, more importantly, in a suitable latent space whose dynamics may control the evolution of their effective dynamics. The challenge here is to identify these latent spaces. Can machine learning extend its undisputed capabilities for spatial pattern recognition to such cases? The automated identification and characterization of these patterns may offer a pathway to solving a wide range of scientific problems and engineering designs with dynamics that span multiple spatio-temporal scales (Wilcox, 1988; Dura-Bernal et al., 2019). Several works in our group have shown that the use of RNN-LSTMS, echo state networks, transformers, and reservoir computing cannot be used to forecast complex systems through data alone (Vlachas et al., 2018, 2020) for times that are large multiples of the respective Lyapunov exponent of system dynamics. Is there a way to use the capabilities of RNN-LSTM and transformers to forecast complex systems for unlimited time frames?

Multiscale methods in scientific computing rely on judicious approximations of the interactions between processes occurring over different scales, and a number of potent frameworks have been proposed. The equation-free framework (EFF) (Kevrekidis et al., 2004) pioneered the notion of alternating between coarse and fine descriptions to efficiently capture the evolution of dynamical systems. In these descriptions, success depends on the separation of scales in the system dynamics and their capability to capture the transfer of information between scales. Perhaps the most important aspect is identifying a suitable latent space (Wiewel et al., 2019). Can ML methods identify and evolve these latent spaces? There have been several efforts to answer this question (for example (Lusch et al., 2018), (Regazzoni et al., 2019), (Lee et al., 2020), (Maulik et al., 2021), (Khoo et al., 2021), (Simpson et al., 2021), (Floryan and Graham, 2022), and (Seyyedi et al., 2023) for a recent survey). Several of these efforts are distinguished by developing a ROM in space, followed by tracking its evolution in time reminiscent of the classical methods for solving PDEs by assuming solutions that are a product of functions of space and time. The effective dynamics learning algorithm (LED) belongs to this category of methods (Vlachas et al., 2022). LED extended the EFF by deploying probabilistic and variational autoencoders (AE) to transfer information between coarse- and fine-scale descriptions and recurring neural networks (RNN) with long-short-term memory (LSTM) (Hochreiter, 1997) gating that evolves the coarse-grained (latent) dynamics. The development of novel attention-based architectures offers new venues for greater expressiveness and specialized mechanisms in computational models. Originally designed for speech and NLP applications (Bahdanau, 2014; Devlin et al., 2019; Radford et al., 2019), these architectures, particularly transformers (Vaswani et al., 2017), are now being adapted for other domains, including dynamical systems (Lee et al., 2019; Fenton et al., 2020). The use of memory in the dynamics of evolving latent spaces is critical, as projection to this space may have distorted the Markovian character associated with PDEs. This salient but important feature of ML versus classical scientific computing should not be overlooked. Transformers, which are made up of encoder-decoder modules with attention mechanisms, excel in handling sequential data, making them promising for time series analysis. However, their quadratic complexity limits scalability for multiscale systems (Baldi, 2021). Recent efforts address the theoretical foundation for attention mechanisms, advancing their understanding and computational capabilities (Baldi and Vershynin, 2023). Transformers can be integrated into LEDs and, as mentioned below, may play a critical role in the expansion of AI generative advances to the prediction of complex systems (Gao et al., 2024a).

LED was able to accurately simulate the dynamics of flow past a cylinder at $Re = 100$ using only two degrees of freedom in the latent space and one degree of freedom in evolving small molecular systems (Vlachas et al., 2021). However, the LED did not predict the flow dynamics at $Re = 1000$ with the same accuracy, exhibiting higher errors in the vicinity of the cylinder surface, indicating that the vorticity generation mechanisms were not captured. At the same time, this failure may indicate ways of combining domain knowledge with the building of suitable encoders. Machine learning algorithms shall be consistent with the underlying physics not simply by adding a term in the cost function, but by designing learning algorithms that can identify causal relationships.

I remark that the EFF framework introduced one more concept for the simulation of multiscale systems: the separate treatment of spatial and temporal dynamics (first coarsen in space, evolve in time, then refine, and evolve). This situation may be reminiscent of the classical method of solving PDEs by the separation of variables. However, this separation of variables has intrinsic limitations and does not account for the coupling between scales in the system dynamics. A possible solution, building on key features of the LED framework, is the adoption of generative AI to couple the space and time evolution of the system dynamics. In the Generative Learning of Effective Dynamics (G-LED) framework (Gao et al., 2024a), instances of high-dimensional data are down-sampled to a lower-dimensional manifold that is evolved through a multi-head, autoregressive attention model, leveraging its low memory footprint and expressivity. In turn, Bayesian diffusion models that map this low-dimensional manifold onto its corresponding high-dimensional space capture the statistics of the system dynamics. G-LED uses a Bayesian diffusion model as a decoder, incorporating physical information through conditional diffusion and virtual observables. The reverse diffusion process flexibly captures the statistics of fields governed by

PDEs (Gao et al., 2024b). It is important to note that the G-LED sequence of snapshots is correlated with the underlying physical process through macro-sequences. Moreover, G-LED decodes multiple consecutive macro-states together as a batch to enhance temporal coherence and increase temporal smoothness in the results. The decoding of multiple macrostates that are contextually connected is reminiscent of OpenAI's Sora, a text-to-video generative AI model that can generate videos of realistic or imaginative scenes from text prompts (Liu et al., 2024).

The exploration of generative AI for solving PDEs is in its infancy. This is a fertile area for a wide range of explorations that can lead to new ways of interfacing AI and scientific computing. The need for plurality in the interface of AI and scientific computing is captured by Jürgen Schmidhuber, who wrote in his pioneering diploma thesis in 1987 (Schmidhuber, 1987) that "... we cannot capture the essence of learning by relying on a small number of algorithms. In contrast, there is a need for a host of context-dependent learning strategies to acquire domain-specific information using information that is already available. Due to the complexity and richness of these strategies and their triggering conditions, the obvious escape seems to be the following. Lastly, give the system the ability to learn how to learn, too. A system with such meta-learning capabilities should view every problem as consisting of at least two problems: Solving it and improving the strategies employed to solve it. Of course, we do not want to stop at the first meta-level!"

4. Discover: the best of both worlds

4.1. Solving forward and inverse problems for PDEs using machine learning

The foundations of scientific computing have been laid on the development of numerical methods for solving PDEs and the effective deployment of these numerical methods in supercomputers. In particular, the field of fluid dynamics and the numerical solution of the Navier–Stokes equations have driven innovations applied to disciplines ranging from epidemics to wildfires. The Navier-Stokes equations involve many classes of PDEs that we distinguish as parabolic, elliptic, and hyperbolic. As mentioned in the abstract, learning numerical solutions of PDEs is a prominent example of inductive learning that is transferable and generalizable. It is a sort of tragedy that today, the curriculum of many Universities no longer includes related classes. The development of numerical discretizations that are consistent with the PDE, while being stable and accurate, has driven the development (and the complexity) of these numerical methods. These methods have been used broadly for solving forward simulations and inverse problems. As already mentioned, scientific computing has limitations on the scales that its methods can resolve. Can ML algorithms replace numerical methods for solving PDEs?

An answer to this question comes from the so-called Physics-Informed ML methods that employ a NN for the approximation of the unknown field. The term "Physics informed" implies employing a cost function in terms of the PDE that, when minimized, produces the weights of the NN and, as such, the solution field. This approach was pioneered for forward problems in PDEs thirty years ago by Lagaris et al. (1998). Recently, the method was revived by Raissi et al. (2019), who used modern ML tools to improve its performance and popularize it as Physics-Informed NN (PINN). The tens of thousands of citations in this paper are a manifestation of interest in developing simple solutions to PDEs without the hassle associated with classical numerical methods. However, there is no free lunch. Although there have been successful demonstrations of PINNs in several benchmark problems (usually 1D in space and time or steady 2D problems), concerns regarding their training and the associated computational cost (Rathore et al., 2024). In PINNs, the cost of evaluating the solution at one point is proportional to the number of weights of the NN. More importantly, the NN approximation may not be consistent with the character of the original differential problem. At least for low-dimensional fields, PINNs are not a viable alternative for solving forward problems. However, there is merit in their use for inverse problems, in particular, as they can easily blend data and equations in formulating the cost function to be minimized. At the same time, there are questions on the well-posedness of such inverse problems and the computational cost as second-order methods requiring a Hessian can be difficult to implement due to the resulting dense Hessians the

applicability of efficient optimization methods such as Newton's method. An extensive study (McGreivy and Hakim, 2024) revealed issues of weak baselines and reporting biases when comparing physics-informed machine learning algorithms with classical numerical methods. Nevertheless, the discussion on physics-informed/enhanced/constrained methods is at its infancy, and one may expect new advances as they serve as convergence points for researchers from different disciplines (Haywood-Alexander et al., 2024). A valuable aspect of physics informed ML methods (He et al., 2020; Wang et al., 2020; Karniadakis et al., 2021). for problems where either the PDEs have missing parameters or no sufficient data is available to form a correct initial-value problem. Such problems are encountered in various fields of science and engineering, and they are handled by various methods such as PDE-constrained optimization (Gunzburger, 2002), data assimilation (Lewis et al., 2006), and system identification (Ljung, 1999). Possible extensions of PINNs to high-dimensional problems in these fields could be very valuable. Research in our group has been inspired by the revival of PINNs. While we had been aware of the work of Lagaris and his coworkers (Lagaris et al., 1998), it was the paper by the Karniadakis group (Cai et al., 2021) that sparked our idea of combining the discretized form of the equations and data to solve inverse problems. However, instead of using NNs to represent the solution, we introduce the discrete form of the equations in the loss function. Minimization of this loss function produces the solution of the equation at the discretization points, a method we call Optimizing the Discrete Loss (ODIL) (Karnakov et al., 2023; Karnakov et al., 2024). ODIL targets inverse problems and combines discrete formulations of PDEs with modern ML tools. The former capability builds on decade-long efforts in numerical analysis and guarantees consistent discretizations. The latter involves automatic differentiation tools such as JAX to allow for flexible software development. ODIL is also related to a number of other approaches that were developed before the recent access to massive data. Solving the discrete equations as a minimization problem is known as the discretize-then-differentiate approach in PDE-constrained optimization (Gunzburger, 2002), and it has been formulated for linear problems as the penalty method (van Leeuwen and Herrmann, 2015). ODIL is also related to the 4D-VAR problem in data assimilation (Lewis et al., 2006). ODIL differs from these methods as its sparse linearization is constructed using automatic differentiation tools, and it is geared towards problems with noisy and gappy data. ODIL has two key components: (i) the discretization that defines the accuracy, stability, and consistency of the method and (ii) the optimization algorithm to solve the discrete problem. In case the underlying problem is sparse, this sparsity is preserved, and the optimization can use a Hessian and achieve a quadratic convergence rate. This rate remains out of reach for stochastic gradient-based training of NNs (Bottou et al., 2018). Moreover, the use of automatic differentiation to compute the Hessian makes the implementation as convenient as applying gradient-based methods. A number of comparisons of the computational cost of ODIL and PINN for a number of benchmark problems has shown (Karnakov et al., 2022) that for the same number of parameters, the methods have comparable accuracy, but ODIL is up to five orders of magnitude faster. A notable application of ODIL has been its recent extension to forecasting the evolution as well as the initiation of gliomas in brain tumors (GLI-ODIL) using real patient multimodal data (Balcerak et al., 2025).

However, I believe that physics-informed machine learning tries to impose onto machine learning principles of scientific computing. A fresh look that may completely bypass PDEs and the usual scientific computing ideas may be a much more powerful approach, as we have already seen in successful applications of data-driven teaching techniques to weather forecasting and ocean dynamics (Kochkov et al., 2024; Price et al., 2024).

4.2. Machine learning for closures of under-resolved PDEs

The development of closures for Reduced Order Models (ROMs), such as coarse-grained or under-resolved partial differential equations (PDEs), is of tremendous importance for fields ranging from aircraft design to weather forecasting (Moser, 2023). In fluid mechanics, closures for large eddy simulation (LES) and Reynolds-averaged Navier–Stokes equations (RANS) have traditionally been developed using physical insight and engineering intuition (Jimenez and Moser, 2000). The vast majority of turbulence

models that use ML are based on Supervised Learning (SL). However, there are lingering questions regarding the generalization of these models beyond the training data (Hickel et al., 2004; Gamahara and Hattori, 2017; Maulik and San, 2017; Vollant et al., 2017; Duraisamy et al., 2019; Fukami et al., 2019; Xie et al., 2019). In SL, neural network parameters (NN) are commonly derived by stochastic gradient descent to minimize the model prediction error. As the error is required to be differentiable with respect to the model parameters, and due to the computational challenge of obtaining chain derivatives through complex, large-scale solvers, SL approaches often define one-step target values for the model (e.g., subgrid scale[SGS] stresses computed from filtered DNS). Due to the single-step cost function in SL, the NN model is not trained to compensate for the systematic discrepancies between DNS and LES and the compounding of numerical discretization errors (Nadiga and Livescu, 2007; Wu et al., 2018; Beck et al., 2019). One way to resolve this difficulty is through the iterative algorithm of Multi-Agent Deep Reinforcement Learning (MADRL). Our group pioneered MADRL as a framework for the systematic construction of such closures (Novati et al., 2021). The key idea of the scientific MADRL method is to simultaneously treat points in the grid as agents that learn to correct discretization errors while considering external cost functions and constraints. Unlike SL, RL optimizes a parametric model by directly exploring the underlying task. In addition, the performance of the RL strategy is measured not by a differentiable objective function but by a cumulative reward. These features are especially beneficial in turbulence modeling, as they allow for avoiding the distinction between a priori and a posteriori evaluations. In the case of LES, the performance of the RL is measured by comparing the statistical properties of the simulation to those of the reference data. MADRL can be trained with limited data as it does not require knowledge of the fully resolved flow field but rather global quantities such as energy spectra. Rather than perfectly recovering SGS computed from filtered simulations that may produce numerically unstable LES (Nadiga and Livescu, 2007), RL can develop novel models that are optimized to reproduce QoI accurately. In addition, automated discovery promoted by the RL shifts the focus from specific models and their parameters to the exploration of spatio-temporal patterns that are inherent to turbulent flows and can allow the generalization of the learned models.

We have obtained state-of-the-art results for homogeneous turbulence (Novati et al., 2021) and wall-bounded turbulent flows (Bae and Koumoutsakos, 2022) that outperform established dynamic SGS modeling approaches. A key part of this success has been the capability to perform policy optimization with the Remember and Forget Experience Replay (ReF-ER) (Novati and Koumoutsakos, 2019) that incorporates efficient sampling has shown state-of-the-art performance on benchmark problems, and can even surpass optimal control algorithms (Novati et al., 2019). MADRL develops the SGS model as a policy that relates agent observations and actions. The learning agents are distributed among the discretization grid points and minimize the discrepancies between the energy spectrum based on the LES and that calculated from fully resolved simulations (DNS). We emphasize that MADRL does not require DNS simulations for its training but can use global quantities such as energy spectra available from experiments or observations. MADRL maximizes high-level objectives and produces SGS models that are stable under perturbation and resistant to compounding errors. Moreover, it offers new paths to solve many of the classic challenges of LES, such as wall-layer modeling, which are difficult to formulate in terms of SL. We believe that this property of MADRL is well suited to wind-wave interaction problems that are faced with relatively limited data. (Cranmer et al., 2020). There is great potential for the interpretability of MADRL through the analysis of effective policies using symbolic computations. Such an analysis involves observation-action pairs that can guide the identification of causal processes in turbulent energy dissipation and the distillation of mechanistic models for multiphysics simulations. Finally, MARL closures enable systematic ablation studies, in terms of the quantities, actions, and rewards of the observed system, that can be used to extract causal information about the processes that determine complex flow dynamics. We argue that, unlike neural network-based approaches, MARL allows for interpretable learning and identification of causal relationships for extreme events. We already have evidence of this capability in 2D oceanic flow simulations (Mojgani et al., 2023). In these simulations, sciMARL was trained to capture the enstrophy spectrum, but it also captured extreme events in terms of the tails of the probability distribution of the vorticity field. This was not achievable with

classical Smagorinsky models. The intervention character of MADRL may be a gateway to explainable ML (Roscher et al., 2020) and effective closures for PDEs. MADRL is a novel, revolutionary strategy for automating the derivation of closures for multi-fidelity ROMs using sparse data.

5. Summary

The speed, veracity, and reliability of simulations for complex systems have a great impact on science and society. Traditionally, PDEs have been the main modeling tool for describing such systems. Simulations based on the solutions of such PDEs have led to many breakthroughs, but they are facing major limitations. Machine learning and data-driven methodologies present new opportunities and frontiers. I suggest that developing algorithmic alloys between AI and scientific computing offers exciting new tools to tackle complex problems, ranging from weather forecasting to epidemics modeling. The focus and cultures of the two disciplines are highly complementary. Where scientific computing is based on numerical analysis, AI is based on modules and architectures; precision is complemented by statistics, and scientific knowledge can be fused with goal-oriented discoveries. In addition to these methods, I believe that exchanges between the two cultures can be beneficial. The scientific computing community can draw a major lesson from the openness in exchanging ideas and software that have empowered the AI and Data Science communities (Donoho, 2024). At the same time, the rigor of scientific computing testing and checks can greatly benefit the rapid evaluation of new ideas in AI (Koumoutsakos, 2024). The clear identification of strengths and weaknesses in each field is a powerful tool to advance science.

We live in very exciting times, with advances in computing and artificial intelligence in all fields of science. One of the greatest contributions of AI has been its ability to capture the imagination of scientists of different disciplines, offering a medium for exchanging ideas. More importantly, it has captured the interest and imagination of a new generation of scientists. It is exciting to see exchanges of ideas between scientists in disciplines such as computer graphics, fluid dynamics, archeology, and psychology. Understanding the world through machine learning models that interact with and challenge those built around PDEs offers new perspectives and new scientific frontiers. We live in very exciting times!

Data availability statement. This is not applicable to this article, as no new data was created or analyzed in this study.

Acknowledgments. I have had the privilege to interact and learn over the last few years from numerous insightful discussions with Dr. Lucas Amoudruz, George Arampatzis, Han Gao, Petr Karnakov, Sebastian Kaltenbach, and Sergey Litvinov. I am grateful to the Swiss Supercomputing Center (CSCS) and, in particular, Dr. Maria Grazia Giuffreda for their unwavering support that has made our research possible (and enjoyable) for over 25 years. Last but not least, I wish to express my gratitude to Professor Eleni Chatzi (ETHZ) for her kindness, encouragement, and patience with me in writing this perspective.

Author contribution. Conceptualization: P.K.; Investigation: P.K.; Resources: P.K.; Writing – original draft: P.K.; Writing – review & editing: P.K.

Funding statement. I am grateful for funding from the US National Science Foundation, DARPA, AFOSR, and the European Research Council.

Competing interests. The author declares none.

References

- Bae HJ and Koumoutsakos P** (2022) Scientific multi-agent reinforcement learning for wall-models of turbulent flows. *Nature Communications* 13(1), 1–9.
- Bahdanau D** (2014) Neural machine translation by jointly learning to align and translate. Preprint, [arXiv:1409.0473](https://arxiv.org/abs/1409.0473).
- Balcerak M, Ezhov I, Karnakov P, Litvinov S, Koumoutsakos P, Weidner J, Zhang RZ, Lowengrub JS, Wiestler B and Menze B** (2025) Individualizing glioma radiotherapy planning by optimization of a data and physics informed discrete loss. *Nature Communications* (accepted).
- Baldi P** (2021) *Deep Learning in Science*. Cambridge: Cambridge University Press.
- Baldi P and Hornik K** (1989) Neural networks and principal component analysis: Learning from examples without local minima. *Neural Networks* 2(1), 53–58.

- Baldi P, Sadowski P and Whiteson D** (2014) Searching for exotic particles in high-energy physics with deep learning. *Nature Communications* 5, 4308.
- Baldi P and Vershynin R** (2023) The quarks of attention: Structure and capacity of neural attention building blocks. *Artificial Intelligence* 319, 103901.
- Bar-Sinai Y, Hoyer S, Hickey J and Brenner MP** (2019) Learning data-driven discretizations for partial differential equations. *Proceedings of the National Academy of Sciences* 116(31), 15344–15349.
- Beck A, Flad D and Munz C** (2019) Deep neural networks for data-driven les closure models. *Journal of Computational Physics* 398, 108910.
- Berkooz G, Holmes P and Lumley JL** (1993) The proper orthogonal decomposition in the analysis of turbulent flows. *Annual Review of Fluid Mechanics* 25(1), 539–575.
- Bodnar C, Bruinsma WP, Lucic A, Stanley M, Brandstetter J, Garvan P, Riechert M, Weyn J, Dong H, Vaughan A, Gupta JK, Thambiratnam K, Archibald AT, Wu C-C, Heider E, Welling M, Turner RE and Perdikaris P** (2024) Aurora: A foundation model of the atmosphere. Preprint, [arXiv:2405.13063](https://arxiv.org/abs/2405.13063).
- Bommasani R, Hudson DA, Adeli E, Altman R, Arora S, von Arx S, Bernstein MS, Bohg J, Bosselut A, Brunskill E, et al.** (2021) On the opportunities and risks of foundation models. Preprint, [arXiv:2108.07258](https://arxiv.org/abs/2108.07258).
- Bottou L, Curtis FE and Nocedal J** (2018) Optimization methods for large-scale machine learning. *SIAM Review* 60(2), 223–311.
- Brunton SL, Noack BR and Koumoutsakos P** (2020) Machine learning for fluid mechanics. *Annual Review of Fluid Mechanics* 52(1), 477–508.
- Buehler MJ** (2024) Mechgpt, a language-based strategy for mechanics and materials modeling that connects knowledge across scales, disciplines, and modalities. *Applied Mechanics Reviews* 76(2), 021001.
- Cai S, Wang Z, Fuest F, Jeon YJ, Gray C and Karniadakis GE** (2021) Flow over an espresso cup: Inferring 3-d velocity and pressure fields from tomographic background oriented schlieren via physics-informed neural networks. *Journal of Fluid Mechanics* 915, A102.
- Carleo G, Cirac I, Cranmer K, Daudet L, Schuld M, Tishby N, Vogt-Maranto L and Zdeborová L** (2019) Machine learning and the physical sciences. *Reviews of Modern Physics* 91(4), 045002.
- Cranmer M, Gonzalez AS, Battaglia P, Xu R, Cranmer K, Spergel D and Ho S** (2020) Discovering symbolic models from deep learning with inductive biases. *Advances in Neural Information Processing Systems* 33, 17429–17442.
- Devlin J, Chang M, Lee K and Toutanova K** (2019) BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. Minneapolis, MN: Association for Computational Linguistics, pp. 4171–4186.
- Donoho D** (2024) Data science at the singularity. *Harvard Data Science Review* 6(1).
- Dura-Bernal S, Suter BA, Gleeson P, Cantarelli M, Quintana A, Rodriguez F, Kedziora David J, Chadderdon GL, Kerr CC, Neymotin SA, McDougal RA, Hines M, Shepherd GM and Lytton WW** (2019) Netpyne, a tool for data-driven multiscale modeling of brain circuits. *Elife* 8, e44494.
- Duraisamy K, Iaccarino G and Xiao H** (2019) Turbulence modeling in the age of data. *Annual Review of Fluid Mechanics* 51, 357–377.
- Fenton M, Shmakov A, Ho T, Hsu S, Whiteson D and Baldi P** (2020) Permutationless many-jet event reconstruction with symmetry preserving attention networks. *Physical Review D*. In press. Also [arXiv:2010.09206](https://arxiv.org/abs/2010.09206).
- Ferrari A and Willcox K** (2024) Digital twins in mechanical and aerospace engineering. *Nature Computational Science* 4(3), 178–183.
- Floryan D and Graham MD** (2022) Data-driven discovery of intrinsic dynamics. *Nature Machine Intelligence* 4(12), 1113–1120.
- Fukami K, Fukagata K and Taira K** (2019) Super-resolution reconstruction of turbulent flows with machine learning. *Journal of Fluid Mechanics* 870, 106–120.
- Gal Y, Koumoutsakos P, Lanusse F, Louppe G and Papadimitriou C** (2022) Bayesian uncertainty quantification for machine-learned models in physics. *Nature Reviews Physics* 4(9), 573–577.
- Gamahara M and Hattori Y** (2017) Searching for turbulence models by artificial neural network. *Physical Review Fluids* 2(5), 054604.
- Gao H, Han X, Fan X, Sun L, Liu L, Duan L and Wang J** (2024b) Bayesian conditional diffusion models for versatile spatiotemporal turbulence generation. *Computer Methods in Applied Mechanics and Engineering* 427, 117023.
- Gao H, Kaltenbach S and Koumoutsakos P** (2024a) Generative learning for forecasting the dynamics of high-dimensional complex systems. *Nature Communications* 15(1), 8904.
- Glaesgen E and Stargel D** (2012) The digital twin paradigm for future NASA and US Air Force vehicles. In *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference 20th AIAA/ASME/AHS Adaptive Structures Conference 14th AIAA*. American Institute of Aeronautics and Astronautics, p. 1818.
- Gunzburger MD** (2002) *Perspectives in Flow Control and Optimization*. Society for Industrial and Applied Mathematics. <https://doi.org/10.1137/1.9780898718720>.
- Haywood-Alexander M, Liu W, Bacska K, Lai Z and Chatzi E** (2024) Discussing the spectrum of physics-enhanced machine learning: a survey on structural mechanics applications. *Data-Centric Engineering* 5, e30.

- He Q, Barajas-Solano D, Tartakovsky G and Tartakovsky AM** (2020) Physics-informed neural networks for multiphysics data assimilation with application to subsurface transport. *Advances in Water Resources* 141, 103610.
- Hickel S, Franz S, Adams NA and Koumoutsakos P** (2004) Optimization of an implicit subgrid-scale model for les. In *Proceedings of the 21st International Congress of Theoretical and Applied Mechanics, Warsaw, Poland*.
- Hochreiter S** (1997) *Long Short-Term Memory: Neural Computation*. Cambridge, MA: MIT-Press.
- Ivanenko AG and Lapa VG** (1967) *Cybernetics and Forecasting Techniques*. New York: American Elsevier Publishing Company.
- Jacobsen C, Zhuang Y and Duraisamy K** (2023) Cocogen: Physically-consistent and conditioned score-based generative models for forward and inverse problems. Preprint, [arXiv:2312.10527](https://arxiv.org/abs/2312.10527).
- Jaynes ET** (2003) *Probability Theory: The Logic of Science*. Cambridge, UK/New York, NY: Cambridge University Press.
- Jimenez J and Moser RD** (2000) Large-eddy simulations: Where are we and what can we expect? *AIAA Journal* 38(4), 605–612.
- Karnakov P, Litvinov S and Koumoutsakos P** (2022) Optimizing a discrete loss (odil) to solve forward and inverse problems for partial differential equations using machine learning tools. Preprint, [arXiv:2205.04611](https://arxiv.org/abs/2205.04611).
- Karnakov P, Litvinov S and Koumoutsakos P** (2023) Flow reconstruction by multiresolution optimization of a discrete loss with automatic differentiation. *The European Physical Journal E* 46(7), 59.
- Karnakov P, Litvinov S and Koumoutsakos P** (2024) Solving inverse problems in physics by optimizing a discrete loss: Fast and accurate learning without neural networks. *PNAS Nexus* 3(1), pgae005.
- Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S and Yang L** (2021) Physics-informed machine learning. *Nature Reviews Physics* 3(6), 422–440.
- Kevrekidis IG, Gear CW and Hummer G** (2004) Equation-free: The computer-aided analysis of complex multiscale systems. *AIChE Journal* 50(7), 1346–1355.
- Kho0 Y, Lu J and Ying L** (2021) Solving parametric pde problems with artificial neural networks. *European Journal of Applied Mathematics* 32(3), 421–435.
- Kochkov D, Yuval J, Langmore I, Norgaard P, Smith J, Mooers G, Klöwer M, Lottes J, Rasp S, Düben P, Hatfield S, Battaglia P, Sanchez-Gonzalez A, Willson M, Brenner MP and Hoyer S** (2024) Neural general circulation models for weather and climate. *Nature* 632(8027), 1060–1066.
- Koumoutsakos P** (2024) On roads less travelled between AI and computational science. *Nature Reviews Physics* 6(6), 342–344.
- Krizhevsky A, Sutskever I and Hinton GE** (2012) Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems* 25, 1097–1105.
- Kurth T, Treichler S, Romero J, Mudigonda M, Luehr N, Phillips E, Mahesh A, Matheson M, Deslippe J, Fatica M, Prabhat P and Houston M** (2018) Exascale deep learning for climate analytics. In *SC18: International Conference for High Performance Computing, Networking, Storage and Analysis*. IEEE, pp. 649–660.
- Lagaris IE, Likas A and Fotiadis DI** (1998) Artificial neural networks for solving ordinary and partial differential equations. *IEEE Transactions on Neural Networks* 9(5), 987–1000.
- Lam R, Sanchez-Gonzalez A, Willson M, Wirsberger P, Fortunato M, Alet F, Ravuri S, Ewalds T, Eaton-Rosen Z, Hu W, Merose A, Hoyer S, Holland G, Vinyals O, Stott J, Pritzel A, Mohamed S and Battaglia P** (2023) Learning skillful medium-range global weather forecasting. *Science* 382(6677), 1416–1421.
- Lee S, Kooshkbaghi M, Spiliotis K, Siettos CI and Kevrekidis IG** (2020) Coarse-scale pdes from fine-scale observations via machine learning. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 30(1), 013141.
- Lee J, Lee Y, Kim J, Kosiorek A, Choi S and Teh YW** (2019) Set transformer: A framework for attention-based permutation-invariant neural networks. In *International Conference on Machine Learning*. PMLR, pp. 3744–3753.
- Lewis JM, Lakshminarayanan S and Dhall S** (2006) *Dynamic Data Assimilation: A Least Squares Approach*, volume 13. Cambridge: Cambridge University Press.
- Li L, Carver R, Lopez-Gomez I, Sha F and Anderson J** (2024) Generative emulation of weather forecast ensembles with diffusion models. *Science Advances* 10(13), eadk4489.
- Li Z, Kovachki N, Azizzadenesheli K, Liu B, Bhattacharya K, Stuart A and Anandkumar A** (2020) Neural operator: Graph kernel network for partial differential equations. Preprint, [arXiv:2003.03485](https://arxiv.org/abs/2003.03485).
- Liu Y, Zhang K, Li Y, Yan Z, Gao C, Chen R, Yuan Z, Huang Y, Sun H, Gao J, He L and Sun L** (2024) Sora: A review on background, technology, limitations, and opportunities of large vision models. Preprint, [arXiv:2402.17177](https://arxiv.org/abs/2402.17177).
- Ljung L** (1999) *System Identification (2nd Ed.): Theory for the User*. USA: Prentice Hall PTR.
- Lusch B, Kutz JN and Brunton SL** (2018) Deep learning for universal linear embeddings of nonlinear dynamics. *Nature Communications* 9(1), 4950.
- Maulik R, Botsas T, Ramachandra N, Mason LR and Pan I** (2021) Latent-space time evolution of non-intrusive reduced-order models using gaussian process emulation. *Physica D: Nonlinear Phenomena* 416, 132797.
- Maulik R and San O** (2017) A neural network approach for the blind deconvolution of turbulent flows. *Journal of Fluid Mechanics* 831, 151–181.
- McGreivyn N and Hakim A** (2024) Weak baselines and reporting biases lead to overoptimism in machine learning for fluid-related partial differential equations. *Nature Machine Intelligence*, 1–14.
- Meng X, Yan X, Zhang K, Da L, Cui X, Yang Y, Zhang M, Cao C, Wang J, Wang X, Gao J, Wang YG, Ji JM, Qiu Z, Li M, Qian C, Guo T, Ma S, Wang Z, Guo Z, Lei Y, Shao C, Wang W, Fan H and Tang YD** (2024) The application of large language models in medicine: A scoping review. *Iscience* 27(5), 109713.

- Milano M and Koumoutsakos P** (2002) Neural network modeling for near wall turbulent flow. *Journal of Computational Physics* 182(1), 1–26.
- Mojgani R, Waelchli D, Guan Y, Koumoutsakos P and Hassanzadeh P** (2023) Extreme event prediction with multi-agent reinforcement learning-based parametrization of atmospheric and oceanic turbulence. Preprint, [arXiv:2312.00907](https://arxiv.org/abs/2312.00907).
- Moser RD** (2023) Numerical challenges in turbulence simulation. In *Numerical Methods in Turbulence Simulation*. Elsevier, pp. 1–43.
- Müller S, Milano M and Koumoutsakos P** (1999) Application of machine learning algorithms to flow modeling and optimization. In *CTR Annual Research Briefs*. Stanford University, pp. 169–178.
- Nadiga BT and Livescu D** (2007) Instability of the perfect subgrid model in implicit-filtering large eddy simulation of geostrophic turbulence. *Physical Review E* 75(4), 046303.
- Novati G, de Laroussilhe HL and Koumoutsakos P** (2021) Automating turbulence modelling by multi-agent reinforcement learning. *Nature Machine Intelligence* 3(1), 87–96.
- Novati G and Koumoutsakos P** (2019) Remember and forget for experience replay. In *Proceedings of the 36th International Conference on Machine Learning*.
- Novati G, Mahadevan L and Koumoutsakos P** (2019) Controlled gliding and perching through deep-reinforcement-learning. *Physical Review Fluids* 4(9), 093902.
- Palmer T** (2015) Modelling: Build imprecise supercomputers. *Nature* 526(7571), 32–33.
- Price I, Sanchez-Gonzalez A, Alet F, Andersson TR, El-Kadi A, Masters D, Ewalds T, Stott J, Mohamed S, Battaglia P, Lam R and Willson M** (2024) Probabilistic weather forecasting with machine learning. *Nature*, 1–7.
- Radford A, Wu J, Child R, Luan D, Amodei D and Sutskever I** (2019) Language models are unsupervised multitask learners. *OpenAI Blog* 1(8), 9.
- Raissi M, Perdikaris P and Karniadakis GE** (2019) Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics* 378, 686–707.
- Rathore P, Lei W, Frangella Z, Lu L and Udell M** (2024) Challenges in training pinns: A loss landscape perspective. Preprint, [arXiv:2402.01868](https://arxiv.org/abs/2402.01868).
- Regazzoni F, Dede L and Quarteroni A** (2019) Machine learning for fast and reliable solution of time-dependent differential equations. *Journal of Computational Physics* 397, 108852.
- Roscher R, Bohn B, Duarte MF and Garcke J** (2020) Explainable machine learning for scientific insights and discoveries. *IEEE Access* 8, 42200–42216.
- Schmidhuber J** (1987) Evolutionary principles in self-referential learning. On learning how to learn: The meta-meta... hook. *Diploma thesis, Institut f. Informatik, Tech. Univ. Munich* 1(2), 48.
- Seyyedi A, Bohlouli M and Oskoe SN** (2023) Machine learning and physics: A survey of integrated models. *ACM Computing Surveys* 56(5), 1–33.
- Shalf J** (2020) The future of computing beyond moore's law. *Philosophical Transactions of the Royal Society A* 378(2166), 20190061.
- Simpson T, Dervilis N and Chatzi E** (2021) Machine learning approach to model order reduction of nonlinear systems via autoencoder and lstm networks. *Journal of Engineering Mechanics* 147(10), 04021061.
- Theodoridis S and Koutroumbas K** (2006) *Pattern Recognition*. Amsterdam: Elsevier.
- Tsetlin ML** (1961) On behaviour of finite automata in random medium. *Avtomat. i Telemekh* 22(10), 1345–1354.
- van Leeuwen T and Herrmann FJ** (2015) A penalty method for pde-constrained optimization in inverse problems. *Inverse Problems* 32(1), 015007.
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł and Polosukhin I** (2017) Attention is all you need. *Advances in Neural Information Processing Systems* 30, 5998–6008.
- Venkatesh KP, Raza MM and Kvedar JC** (2022) Health digital twins as tools for precision medicine: Considerations for computation, implementation, and regulation. *npj Digital Medicine* 5(1), 150.
- Vlachas PR, Arampatzis G, Uhler C and Koumoutsakos P** (2022) Multiscale simulations of complex systems by learning their effective dynamics. *Nature Machine Intelligence* 4(4), 359–366.
- Vlachas PR, Byeon W, Wan ZY, Sapsis TP and Koumoutsakos P** (2018) Data-driven forecasting of high-dimensional chaotic systems with long short-term memory networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 474(2213), 20170844.
- Vlachas PR, Pathak J, Hunt BR, Sapsis TP, Girvan M, Ott E and Koumoutsakos P** (2020) Backpropagation algorithms and reservoir computing in recurrent neural networks for the forecasting of complex spatiotemporal dynamics. *Neural Networks* 126, 191–217.
- Vlachas PR, Zavadlav J, Praprotnik M and Koumoutsakos P** (2021) Accelerated simulations of molecular systems through learning of effective dynamics. *Journal of Chemical Theory and Computation* 18(1), 538–549.
- Vollant A, Balarac G and Corre C** (2017) Subgrid-scale scalar flux modelling based on optimal estimation theory and machine-learning procedures. *Journal of Turbulence* 18(9), 854–878.
- Wang R, Kashinath K, Mustafa M, Albert A and Yu R** (2020) Towards physics-informed deep learning for turbulent flow prediction. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1457–1466.

- Wiener N** (1950) Cybernetics. *Bulletin of the American Academy of Arts and Sciences* 3(7), 2–4.
- Wiewel S, Becher M and Thuerey N** (2019) Latent space physics: Towards learning the temporal evolution of fluid flow. In *Computer Graphics Forum*, volume 38, Wiley Online Library, pp. 71–82.
- Wilcox DC** (1988) Multiscale model for turbulent flows. *AIAA Journal* 26(11), 1311–1320.
- Wu J, Xiao H and Paterson E** (2018) Physics-informed machine learning approach for augmenting turbulence models: A comprehensive framework. *Physical Review Fluids* 3(7), 074602.
- Xie C, Wang J, Li H, Wan M and Chen S** (2019) Artificial neural network mixed model for large eddy simulation of compressible isotropic turbulence. *Physics of Fluids* 31(8), 085112.
- Zavadlav J, Arampatzis G and Koumoutsakos P** (2019) Bayesian selection for coarse-grained models of liquid water. *Scientific Reports* 9(1), 99.