
MACHINE LEARNING IN FLUID DYNAMICS: A CRITICAL ASSESSMENT

Kunihiko Taira

Department of Mechanical and Aerospace Engineering
University of California, Los Angeles, California 90095, USA
ktaira@seas.ucla.edu

Georgios Rigas

Department of Aeronautics
Imperial College London, London, SW7 2AZ, UK
g.rigas@imperial.ac.uk

Kai Fukami

Department of Aerospace Engineering
Tohoku University, Sendai, Miyagi 980-8579, Japan
kfukami1@tohoku.ac.jp

August 26, 2025

ABSTRACT

The fluid dynamics community has increasingly adopted machine learning to analyze, model, predict, and control a wide range of flows. These methods offer powerful computational capabilities for regression, compression, and optimization. In some cases, machine learning has even outperformed traditional approaches. However, many fluid mechanics problems remain beyond the reach of current machine learning techniques. As the field moves from its current state toward a more mature paradigm, this article offers a critical assessment of the key challenges that must be addressed. Tackling these technical issues will not only deepen our understanding of flow physics but also expand the applicability of machine learning beyond fundamental research. We also highlight the importance of community-maintained datasets and open-source code repositories to accelerate progress in this area. Furthermore, the future success of machine learning in fluid dynamics will depend on effective training — not only for the next generation of researchers but also for established fluid mechanicians adapting to this evolving landscape. Data-driven fluid dynamics is in its critical transitional state over the next few years to shape its future. This perspective article aims to spark discussions and encourage collaborative efforts to advance the integration of machine learning in fluid dynamics.

1 Introduction

Fluid dynamics has three subfields of experimental, theoretical, and computational fluid dynamics. In the past several years, data-driven fluid dynamics has emerged as a powerful fourth subfield to support the overall endeavor of fluid dynamics by leveraging data from experiments and simulations. The recent level of interest in data-driven techniques by the research community has been staggering, fueled by the enormous advancements in computational hardware and software developments, which especially have enabled the use of deep learning. Indeed, it has been amazing to observe the rapid and widespread growth of data-driven fluid dynamics in recent years [1, 2].

In its early stage of data-driven analysis, most studies involved the application of basic data-science techniques to canonical fluid flow problems to analyze and model flow physics. Such approaches often leveraged linearization and/or space-time decomposition to analyze flows. However, the advent of modern machine learning, particularly its ability to handle non-convex optimization and nonlinear functional approximations, has enabled the incorporation of nonlinear analysis into data-driven frameworks. This advancement allows researchers to capture complex, nonlinear flow physics and correlations that were previously difficult to resolve using traditional methods. These efforts have

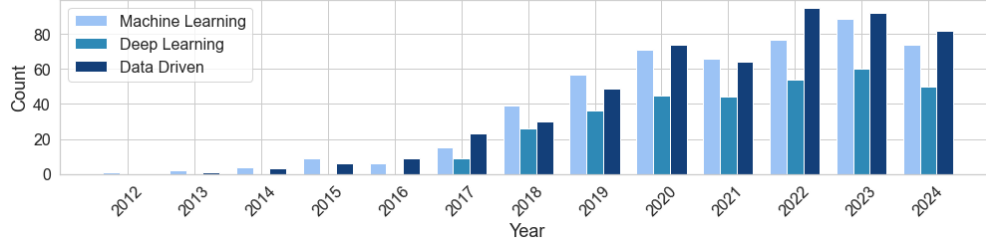


Figure 1: The number of presentations from 2012-2024 APS-DFD meetings that contain “machine learning,” “deep learning,” and “data driven” in their titles or abstracts (search performed on May 1, 2025).

opened new avenues to study, model, and control complex fluid flows without stringent assumptions [3, 4]. This point is a major strength of machine learning and supports the analysis of fluid flows with complex physics.

Modern machine-learning techniques have been used for a range of fluid mechanics problems in recent years [2]. For compression of flow field data, nonlinear machine learning methods can exhibit superb performance compared to traditional approaches [5, 6]. The resulting low-dimensional representation of complex flow fields can be used to model their dynamics and perform control [7, 8, 9]. Finding nonlinear relationships between input and output variables is another strength of machine learning, enabling field estimation [10, 11] and super resolution [12, 13, 14]. Moreover, there are major efforts in using machine learning for the development of turbulence models [15, 16, 17, 18, 19] and optimizing flow control [20], navigation [21, 22], and design strategies [23]. These efforts are only a handful of studies that the fluid dynamics community has seen.

To illustrate recent trends in the use of machine learning in fluid dynamics, we present in figure 1 the number of the American Physical Society Division of Fluid Dynamics (APS-DFD) meeting abstracts that contain the term “data driven,” “machine learning,” or “deep learning”¹. We observe that there has been a growing number of abstracts with machine learning/data-driven techniques being mentioned from 2012 to 2022. It should be noted that there may be additional presentations that likely have discussed data-driven efforts without the aforementioned search words being listed. What is also noteworthy is that from about 2022, there has been a plateau in the number of abstracts using those keywords. We suspect that this plateau reflects the community no longer viewing data-driven approaches as something special, but rather, considering such techniques as part of the regular toolbox to study fluid dynamics. As there is an increasing number of data-driven fluid dynamics workshops and conferences being held all over the globe, the impact of data-driven fluid dynamics is not plateauing but rising.

With the initial ramp-up phase of machine learning in fluid dynamics appearing to stabilize, data-driven fluid dynamics is coming to its transitional state from its initial growth stage to its next stage. In some ways, the data-driven fluid dynamics community is entering the teenage years embarking on its soul-searching journey to determine how the subfield should shape itself as it further establishes its foundation and works together with the other subfields of fluid dynamics. The purpose of this perspective article is to provide a critical assessment of the major outstanding issues in machine learning for fluid dynamics. While we do not attempt to provide a review of the field or to predict future trends in this article, we hope the discussions provided herein stimulate and spark innovative ideas to support future machine learning research and educational activities in fluid dynamics.

2 Discussion

Let us offer a critical assessment of outstanding issues related to the use of machine learning in fluid dynamics. Although these issues may appear as hurdles at first glance, they could be areas of opportunity for future activities. Even if they cannot be immediately addressed, we believe that being mindful of them can support individual studies and help shape community research directions.

2.1 Formulation

In the context of physical applications, including fluid mechanics, *machine learning* refers to the use of data-driven algorithms to uncover patterns, model complex dynamics, and approximate operators or functionals that may be difficult or impossible to express analytically. A machine learning model f_θ is typically trained on simulation or experimental data to capture nonlinear relationships between the input features \mathbf{x}_i (e.g., state variables, spatial coordinates, time,

¹In contrast to these terms, it is interesting to note that the use of “artificial intelligence” has not been prevalent in fluid dynamics.

boundary conditions, or other problem-specific variables) and the output \mathbf{y}_i (e.g., physical quantities composed of traditional or generic observables). The machine learned model can be numerically found by solving a non-convex optimization problem of

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \mathcal{L}(\theta) = \underset{\theta}{\operatorname{argmin}} \left[\frac{1}{N} \sum_{i=1}^N \mathcal{L}_i(f_{\theta}(\mathbf{x}_i), \mathbf{y}_i) + \sum_j \lambda_j \mathcal{R}_j(\theta) \right]. \quad (1)$$

Here, the goal is to find the optimal set of model parameters θ^* that minimizes a loss function $\mathcal{L}(\theta)$. The model parameters θ in modern machine learning are generally high-dimensional (e.g, weights of a neural network) [24, 25, 26].

One of the most critical steps in machine learning is the setup of this mathematical problem. The loss function typically consists of two primary components: a data fidelity term that measures the discrepancy between model predictions $f_{\theta}(\mathbf{x}_i)$ and actual observations \mathbf{y}_i across N training samples, and regularization terms $\mathcal{R}_j(\theta)$ weighted by hyperparameters λ_j to prevent overfitting [26]. The parameter vector θ contains all the adjustable values in the model (e.g., weights and biases inside the neural network). The resulting non-convex optimization landscape is typically navigated using variants of stochastic gradient descent that iteratively update parameters in the direction that reduces the loss function. With this setting, a machine learning model can be used to represent a nonlinear operator, including a flow map, a closure model, or an input-output relation, enabling tasks such as reduced-order modeling, flow prediction, or control [1, 2, 15].

Generally, it is helpful to incorporate physical insights into the formulation of the problem to facilitate the machine-learning process and the interpretation of the results [27, 28]. One way to achieve this is to add loss terms that penalize violations of the governing equations via the residuals of the relevant differential operators [29, 30, 31], as part of the last term in Eq. (1). The pros and cons of utilizing a specific machine learning algorithm and each loss function term should be considered carefully. We caution that over-constraining the training objective is not advisable, as an appropriately compiled training data should allow a properly chosen machine learning algorithm to extract relevant information [32]. One should also be mindful of whether a sufficient amount of training, validation, and test data is available to enable the machine-learning study to provide robust and generalizable findings.

Data-driven analysis should also be performed in steps with respect to the problem complexity. It is often useful to test the desired machine-learning approach first for small-scale, canonical model problems before taking on the full-scale, complex fluid flow problems. Although this workflow may seem obvious, the powerful capability of machine learning often oversteps it. Such simple steps can, however, help with the verification of the approach and gain a deeper understanding of how the approach provides insights into the problem. These steps can also identify the strengths and weaknesses of the selected approach in assessing the overall framework and may uncover possible improvements of the chosen model, mathematical problem setting, and the training dataset.

2.2 Gaining Insights

One of the biggest questions that the data-driven fluid dynamics community faces is whether the use of machine learning is actually deepening our understanding of fluid dynamics. This seemingly simple question requires some thought. First of all, what do we mean by understanding? The term *understanding*, has different meanings depending on whether a person is trying to solve a practical problem or aiming to gain deep insights into a physical phenomenon. These two objectives would lead to different ideas of insights. In the case of solving a practical problem, information that helps solve the problem could be considered valuable insights. However, such insights may not necessarily deepen our knowledge of physics. In contrast, if the objective is to achieve enhanced knowledge of flow physics, we seek to relate the findings from machine learning to physical observations, knowledge, and theory at a deeper level while generalizing beyond the training data sets. The latter objective calls for a more holistic approach and requires working with all four subfields of fluid dynamics.

We should also be cautious about the interpretability of nonlinear machine-learning methods. It was only a few decades ago that proper orthogonal decomposition (POD) was considered an advanced (linear) technique that produced findings that were not interpretable by the fluid dynamics community. However, the community now considers POD [33, 34] as a mathematically rigorous data-driven technique that offers physically and mathematically interpretable results. With the research community becoming familiar with the methods over time, what is considered “interpretable” will evolve as the POD example reminds us. This is also true for modern (nonlinear) machine-learning techniques for which well-experienced data-driven fluid dynamics researchers are now able to extract physical insights from latent variables [8], which do not reside in the physical (ambient) space. Such insights can be gained with proper experience in machine learning and by carefully relating the findings to fluid dynamics. It is anticipated that many results produced from novel machine-learning techniques will be digested in a manner relatable to flow physics in the coming years, if not

already. As such, we expect the meaning of interpretability and the labeling of interpretable methods to actively evolve as the data-driven fluid dynamics field grows and matures in the future.

2.3 Model Generalizability

Given the diverse flow configurations, non-dimensional parameters, and a range of spatiotemporal flow characteristics we encounter in fluid dynamics, the generalizability of machine-learning models is a major challenge. A truly generalizable model should perform well for a range of flow scenarios (e.g., from laminar to turbulent), while exhibiting robustness against variations in data fidelity, spatial resolution, and sensor noise. Machine-learning models are, however, often reliable only within the regime for which testing conditions are similar to those seen in training. We should be mindful that no single model can be optimal for all possible scenarios in the absence of prior knowledge [35]. This suggests a trade-off between model accuracy on a specific task and its ability to generalize across different flow conditions.

The validity of data-driven models is often discussed in the context of interpolation and extrapolation. Modern machine-learning techniques may work outside the training regime if the models are able to capture existing trends across spatial scales or through nonlinear relationships. This may appear as extrapolation in the traditional sense, but it generally amounts to interpolation in the machine learned coordinates. Although caution is strongly advised, it is possible to observe pleasantly surprising performance even outside of the training data if the analysis is conducted appropriately [36]. However, when the machine learning model fails outside the trained region over machine-learned coordinates, it does so spectacularly. Understanding the transitional characteristics between interpolation and extrapolation requires a detailed analysis of the model in a machine-learned space that features the data distribution. This is an open question in data-driven fluid dynamics that requires further examination, especially when identifying and modeling rare events.

For applying trained models to unseen conditions, transfer learning can be useful. This learning technique enables models trained with one particular flow problem to be reused and re-trained faster for another one. Transfer learning has enabled the extension and fine tuning of machine learning models from lower Reynolds number flows for higher Reynolds number flows with reduced training efforts for flow estimation [37], reinforcement learning based flow control [38], and turbulence modeling [39]. Another promising concept to extend the utility of existing models is the foundation model [40, 41]. This approach is often taken in image science, which fixes initial layers of a deep network to learn broadly applicable representations, while only fine-tuning the latter layers near the output for specific downstream tasks. These general concepts may extend our current machine learning models to be applicable to a much wider range of fluid flow problems.

2.4 Model Complexity

Fundamental studies in the past have shown that functions or governing equations that describe physical phenomena are generally composed of simple, compact expressions. In this vein, the best model should be parsimonious in its form for it to be interpretable and generalizable to describe the physics [2, 42, 43, 44]. Basically, one should not use a complicated representation if there is a simple way to describe the phenomenon.

As a general trend in any novel methods being used for fluid mechanics, initial efforts tend to be fairly simple in the way methods are utilized to solve problems. Machine-learning techniques are no exception in how they have been introduced to the fluid dynamics community. In recent years, however, the complexity of the machine-learning techniques being used for some studies has increased at a rate that has not been seen by our past data analysis techniques. This is in part due to the machine-learning techniques' ability to hold a large number of parameters in the hidden layers and algorithms. Tuning such a set of parameters used to be an enormous challenge for non-convex optimization problems (resulting from machine learning). However, it is now becoming significantly easier with the advancement in software and hardware resources. As such, there has been less hesitation for machine-learning practitioners to propose data-driven techniques that are surprisingly and unnecessarily complex, often due to a "let's try and see if it works" mentality. For instance, some efforts have reported the use of machine-learning-based methods that require a number of tunable parameters far larger than what a computational fluid dynamics software would require to accurately simulate the flow to predict even the simplest flow behavior.

For a machine learning model, sparsity-promoting methods are capable of removing terms or network connections that are physically irrelevant or subdominant. In many cases, they encourage interpretability of the returned results. Moreover, when applied to the spatial domain, they can develop locally focused models instead of global ones. These concepts have led to the identification of many sparse dynamical systems models [45] and partial differential equations [46] from flow field data.

Without striving for simple or appropriate representation, machine-learning-based analysis may not be able to break away from being perceived as black-box models that as a conduit between the input and output data. There are uses for such work, but they often would not yield deeper insights into the physics of fluid flows.

2.5 Data Repositories and Code Sharing

As data plays an ever more important role in advancing science and engineering with machine learning, the utility of data should be carefully considered beyond a single or limited research use. This means that training datasets should be compiled with downstream use for a large audience in mind. With machine learning algorithms generally being data-intensive, data as a resource should be shared across the community to advance the field of fluid dynamics. Without such a shared resource, a researcher will be limited to a modest collection of data attainable by an individual research group, restricting machine learning research efforts.

There are three particularly important matters when compiling an archival data set. First, to enable machine learning models to be as generalizable as possible across a range of conditions, the dataset must span a variety of physical phenomena. Second, the dataset must be of high quality, faithfully capturing the intended flow physics without significant error or bias. Third, the distribution of data should be carefully assessed to make sure sampling bias is minimized within an acceptable range [47, 48]. For example (while it may be unavoidable for now), there is a tendency for computational fluid dynamics data to be associated with lower Reynolds numbers compared to experimental measurements.

The fluid flow data repositories hosted by Johns Hopkins University for canonical turbulent flow data sets [49, 50] have supported the development of analysis and modeling of turbulent flows with lasting impact. There are also repositories made available in recent years with unsteady flow data for machine learning applications in mind [51, 52]. As machine learning models can learn flow physics from not only a limited number of cases but a large number of flows across combinations of parameters, we require a much larger data repository comprising a sizable ensemble of cases [52]. This means that such a dataset would have a large number of spatial grid points, temporal snapshots, and flow cases, making a necessary data library to be of a completely different size from what we are accustomed to hosting by a single research group or center.

Consequently, one of the major issues is the hosting of such a data repository. While most research support is provided for research projects and equipment, rarely do we see support for hosting a data server with dedicated staff. This matter needs serious attention to enable researchers to freely share datasets and advance machine learning research efforts. Moreover, it may require a change in our perception to consider flow data repositories as important archival resources that may require libraries or national/international research centers to curate important spatiotemporally resolved flow field data across a range of parameters for various flows. In fact, our community is in need of a discussion to enhance findability, accessibility, interoperability, and reusability (FAIR) [53] of training data sets for machine learning in fluid dynamics. We can learn from other areas of research, such as the astronomy and atmospheric science communities, that have a communal data repository. Once the fluid dynamics community establishes guidelines on data sharing setups and formats, we will be able to further accelerate the research progress in machine learning and develop generalizable and robust models. It may also invite new data formats for analysis beyond probe data and snapshots [54]. Such a data repository will bring benefit to the entire fluid dynamics community beyond the machine learning groups and the scientific community at large.

In support of FAIR, it is important to share the codes for machine-learning analysis in the community. There has been encouraging progress by some journals and research groups, making their machine-learning codes and datasets publicly available through online platforms such as GitHub and Zenodo. Such efforts can enhance not only the reproducibility of the results but also collaboration across the world. To further support this promising trend, it would be necessary to address several remaining challenges, such as a lack of standardization in code structure and documentation, likely causing difficulties in reproducing results and extending the code to their own data sets. In addition to these technical matters, the community needs to culturally shift toward a state of transparent research practices [55].

Let us also mention an area that requires some attention related to the coding effort. There are well-established guidelines on verification, validation, and uncertainty qualification for computational and experimental fluid dynamics [56, 57, 58]. Such procedures ensure that studies performed in these fields provide reliable insights. Unfortunately, for many machine-learning efforts, there does not appear to be a similar level of rigor that establishes the architectures of the models and ensures the fidelity and performance of their findings [59, 28]. This is not to say such processes do not exist in the machine-learning community. However, they are often overlooked. Furthermore, given the complexities of machine learning, it would be ideal to have the ability to estimate its necessary computational and memory resources in a manner analogous to how scaling is performed to determine the required grid resolutions across Reynolds numbers in computational fluid dynamics analysis. This current state of the matter urgently calls for some guidelines to be

developed such that the robustness and fidelity of machine-learning techniques can be guaranteed in future efforts. With proper verification and validation performed and limitations identified, higher credibility can be attributed to machine-learning models.

While sharing resources should be encouraged, we should remember that collecting high-quality data is an endeavor of its own. Data-driven fluid dynamics is reliant on experts in experimental and computational fluid dynamics to collect the data. Proper credit always needs to be given to the data creators/collectors.

2.6 Education

Given the enormous progress and excitement seen in the general machine learning community, it is perhaps understandable to have high expectations that machine learning can serve as a breakthrough technology to solve some of the most challenging problems in fluid dynamics. However, we must have healthy expectations of what can be achieved by machine learning. It is important to recognize that not all developments in general machine learning can be easily transferred to the fluid dynamics community due to the difference in the research objectives and the characteristics of flow physics. Machine learning practitioners in fluid dynamics should also have a deep understanding of the flow physics, as incorporating prior knowledge about the flow can improve the machine learning process and facilitate the extraction of insights from data. For undergraduate and graduate education, a class on data-driven fluid dynamics should be developed in concert with theoretical, computational, and experimental fluid dynamics courses. Furthermore, exposing students to machine learning in their early stages of fluid mechanics education through homework assignments and projects that involve data-driven analysis may strengthen their ability to use such tools to a broader class of problems and prepare them for proper usage of machine learning analysis in fluid dynamics.

There is a new type of problem that our community is facing in terms of the need for machine learning education. As the younger generation enters the research community already with some familiarity with machine learning, it can be challenging even for established researchers to keep up with the younger minds. Many engineering and science students now enter graduate school with machine learning experience. This raises a question of not only how the community should teach and advise the incoming research population, but also how we should train the pre-machine learning research population. This generational gap in machine learning training needs to be filled to establish a balanced perspective on fluid dynamics and maintain healthy expectations on what machine learning can achieve in research and development. For these reasons, educational opportunities should be made available to the wider research population at different levels of familiarity with machine learning.

In response to these points, data-driven fluid dynamics conferences and workshops should involve a diverse set of stakeholders, including those from all four subfields of fluid dynamics working in academia, industry, and government. Moreover, workshops aimed at non-practitioners are equally important as those targeted at spreading the state of the art of machine learning. By doing so, we will be able to achieve the integration of machine learning techniques in various studies of fluid dynamics, when appropriate, and advance the overall field. Essentially, we will be able to blend the four subfields of fluid dynamics seamlessly across generations. Moreover, fundamental and applied machine learning research in fluid dynamics must work hand-in-hand. Only with a fundamental understanding of how machine learning can make a difference in fluid mechanics, the swift transfer of machine learning technology from basic science to practical applications can be enabled [60, 61].

At last, machine learning practitioners in fluid dynamics should have good knowledge of computational and experimental fluid dynamics to understand what constitutes good data. It is also important to appreciate how uncertainties and biases can creep into the training dataset. As machine learning intimately relies on the quality of the training data, the data-driven fluid dynamics community must work closely with the experts who collect high-quality data from simulations and experiments. Blending these subfields of fluid dynamics will advance the reliability of future machine learned models and enhance the insights gained from their learning process.

3 Remarks

Data-driven fluid dynamics has undergone tremendous growth over the past decade and has become recognized as a powerful fourth subfield in fluid dynamics, complementing theoretical, experimental, and computational fluid dynamics. In fact, the usage of data-driven techniques has become so widespread in fluid dynamics research that it is no longer special to see it in conference presentations or archival articles. As the field of data-driven fluid dynamics aims to advance itself further, it requires careful examination of some outstanding issues. This perspective article calls attention to machine learning formulation, gaining insights, generalizing models, considering model complexity, sharing data and codes, and education. It is our sincere hope that this article stimulates future discussions and invites innovative ideas.

There remain many challenging problems in fluid dynamics that involve multi-physics and unknown mechanisms. These problems may not have an established governing equation, for which machine learning may serve as a powerful tool to make headway. It may also perform well for problems with strong nonlinearity that traditional techniques have difficulty analyzing. There are fluid dynamics problems that do not have the luxury of having sufficient data due to challenges associated with conducting a large number of experiments or simulations. Machine learning efforts will have to adapt their techniques to sample and leverage data in a smart manner. In such a case, incorporating prior knowledge about the physics may facilitate the learning process even from a minimal amount of data [62]. Moreover, there is growing interest in applying generative models, including large language models, diffusion models, and other foundation models [63, 64], to fluid mechanics problems. These approaches can employ chain-of-thought reasoning [63] to explicitly trace the logical steps underlying fluid flow analysis and prediction. By making the reasoning process transparent, these techniques offer potential for both deepening our mechanistic understanding of fluid phenomena and improving predictive accuracy across diverse flow regimes [64].

The development of machine learning techniques has been closely tied to hardware advancements. The availability of powerful GPUs has catapulted the recent movement in deep learning. With the GPU architecture enabling large-scale nonlinear optimization problems to be numerically solved, deep learning methods and techniques evolving around them have been enjoying the benefits. The current algorithmic trends will evolve in the future as novel computing architectures elicit innovations in next-generation machine learning approaches, including those associated with quantum computing [65, 66]. As CPUs and GPUs have influenced computational science and machine learning, the future of fluid dynamics will likely see a transformation with advancements in future computing platforms.

Before closing, let us briefly comment on publishing works on data-driven fluid dynamics. The community should maintain the same high standards of publication for manuscripts on the topic of theoretical, experimental, computational, and data-driven fluid dynamics, or combinations thereof. This means that papers in data-driven fluid dynamics should deepen our understanding of flow physics or propose novel methodologies that can aid discoveries in fluid dynamics [67]. For the latter type of work, Physical Review Fluids houses a topical section entitled *Methods: New Experiments, Algorithms, and Theory* (NEAT) [68], which can serve as a home to breakthrough data-driven fluid dynamics papers. As our research community goes through this transitional period for machine learning in fluid dynamics, it is very exciting to imagine the kinds of innovative papers we will see in the coming years.

The vast advancements made by data-driven fluid dynamics have been nothing short of incredible. With the aforementioned outstanding problems addressed in the coming years, one can only imagine how much progress the community will be able to make. The intellectual investments we make in the data-driven fluid dynamics during the current transitional state will form the basis of how machine learning will expand the future horizon of fluid dynamics in concert with theoretical, experimental, and computational efforts.

acknowledgments

We thank Jeff D. Eldredge, Alec J. Linot, Beverley J. McKeon, Douglas R. Smith, and Jonathan Tran for their valuable comments on the draft version of this article. KT acknowledges support from the Vannevar Bush Faculty Fellowship (N00014-22-1-2798), the Air Force Office of Scientific Research (FA9550-21-1-0178, FA2386-25-1-1003), and the Army Research Office (W911NF-24-1-0213). GR acknowledges support from the UKRI AI for Net Zero grant (EP/Y005619/1). KF acknowledges support from the JSPS KAKENHI Grant Number JP25K23418.

References

- [1] M. P. Brenner, J. D. Eldredge, and J. B. Freund. Perspective on machine learning for advancing fluid mechanics. *Phys. Rev. Fluids*, 4:100501, Oct 2019.
- [2] S. L. Brunton, B. R. Noack, and P. Koumoutsakos. Machine learning for fluid mechanics. *Annu. Rev. Fluid Mech.*, 52:477–508, 2020.
- [3] C. Lee, J. Kim, D. Babcock, and R. Goodman. Application of neural networks to turbulence control for drag reduction. *Phys. Fluids*, 9(6):1740–1747, 06 1997.
- [4] M. Milano and P. Koumoutsakos. Neural network modeling for near wall turbulent flow. *J. Comput. Phys.*, 182:1–26, 2002.
- [5] M. A. Kramer. Nonlinear principal component analysis using autoassociative neural networks. *AIChE J.*, 37(2):233–243, 1991.
- [6] K. Fukagata and K. Fukami. Compressing fluid flows with nonlinear machine learning: mode decomposition, latent modeling, and flow control. *Fluid Dyn. Res.*, 57:041401, 2025.

- [7] Alec J. Linot, Kevin Zeng, and Michael D. Graham. Turbulence control in plane Couette flow using low-dimensional neural ODE-based models and deep reinforcement learning. *Int. J. Heat Fluid Flow*, 101:109139, 2023.
- [8] Kai Fukami and Kunihiro Taira. Grasping extreme aerodynamics on a low-dimensional manifold. *Nat. Commun.*, 14:6480, 2023.
- [9] Kai Fukami, Hiroya Nakao, and Kunihiro Taira. Data-driven transient lift attenuation for extreme vortex gust-airfoil interactions. *J. Fluid Mech.*, 992:A17, 2024.
- [10] Krithika Manohar, Bingni W. Brunton, J. Nathan Kutz, and Steven L. Brunton. Data-driven sparse sensor placement for reconstruction: Demonstrating the benefits of exploiting known patterns. *IEEE Contr. Syst. Mag.*, 38(3):63–86, 2018.
- [11] K. Fukami, R. Maulik, N. Ramachandra, K. Fukagata, and K. Taira. Global field reconstruction from sparse sensors with Voronoi tessellation-assisted deep learning. *Nat. Mach. Intell.*, 3:945–951, 2021.
- [12] Kai Fukami, Koji Fukagata, and Kunihiro Taira. Super-resolution reconstruction of turbulent flows with machine learning. *J. Fluid Mech.*, 870:106–120, 2019.
- [13] H. Kim, J. Kim, S. Won, and C. Lee. Unsupervised deep learning for super-resolution reconstruction of turbulence. *J. Fluid Mech.*, 910:A29, 2021.
- [14] S. Maejima and S. Kawai. Unsupervised and supervised machine learning pipeline for super-resolution-based sub-grid scale modeling in coarse-grid large-eddy simulations. *J. Fluid Mech.*, 1013:A28, 2025.
- [15] K. Duraisamy, G. Iaccarino, and H. Xiao. Turbulence modeling in the age of data. *Annu. Rev. Fluid Mech.*, 51(1):357–377, 2019.
- [16] J. Ling, A. Kurawski, and J. Templeton. Reynolds averaged turbulence modelling using deep neural networks with embedded invariance. *J. Fluid Mech.*, 807:155–166, 2016.
- [17] A. Lozano-Durán and H. J. Bae. Machine learning building-block-flow wall model for large-eddy simulation. *J. Fluid Mech.*, 963:A35, 2023.
- [18] M. Oulghelou, S. Cherroud, X. Merle, and P. Cinnella. Machine-learning-assisted blending of data-driven turbulence models. *Flow Turbul. Combust.*, in press, 2025.
- [19] H. Jane Bae and Petros Koumoutsakos. Scientific multi-agent reinforcement learning for wall-models of turbulent flows. *Nat. Commun.*, 13(1):1443, 2022.
- [20] Jean Rabault, Miroslav Kuchta, Atle Jensen, Ulysse Réglade, and Nicolas Cerardi. Artificial neural networks trained through deep reinforcement learning discover control strategies for active flow control. *J. Fluid Mech.*, 865:281–302, 2019.
- [21] Peter Gunnarson, Ioannis Mandralis, Guido Novati, Petros Koumoutsakos, and John O. Dabiri. Learning efficient navigation in vortical flow fields. *Nat. Commun.*, 12(1):7143, 2021.
- [22] Zhecheng Liu, Diederik Beckers, and Jeff D. Eldredge. Model-based reinforcement learning for control of strongly disturbed unsteady aerodynamic flows. *AIAA J.*, in press(in press):1–21, 2025.
- [23] Andrew Mole, Max Weissenbacher, Georgios Rigas, and Sylvain Laizet. Reinforcement learning increases wind farm power production by enabling closed-loop collaborative control. arXiv:2506.20554, 2025.
- [24] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [25] Jeremy Watt, Reza Borhani, and Aggelos K. Katsaggelos. *Machine Learning Refined: Foundations, Algorithms, and Applications*. Cambridge University Press, 2 edition, 2020.
- [26] Steven L. Brunton and J. Nathan Kutz. *Data-Driven Science and Engineering: Machine Learning, Dynamical Systems, and Control*. Cambridge University Press, 2 edition, 2022.
- [27] Steven L. Brunton. Applying machine learning to study fluid mechanics. *Acta Mech. Sin.*, 37(12):1718–1726, 2021.
- [28] Petros Koumoutsakos. Machine learning and partial differential equations: benchmark, simplify, and discover. *Data-Centric Eng.*, 6:e29, 2025.
- [29] M. Raissi, P. Perdikaris, and G.E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *J. Comput. Phys.*, 378:686–707, 2019.
- [30] George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nat. Rev. Phys.*, 3(6):422–440, 2021.

- [31] Y. Patel, V. Mons, O. Marquet, and G. Rigas. Turbulence model augmented physics-informed neural networks for mean-flow reconstruction. *Phys. Rev. Fluids*, 9(3):034605, 2024.
- [32] R. Sutton. The bitter lesson, 2019.
- [33] G. Berkooz, P. Holmes, and J. L. Lumley. The proper orthogonal decomposition in the analysis of turbulent flows. *Annu. Rev. Fluid Mech.*, 25:539–575, 1993.
- [34] K. Taira, S. L. Brunton, S. Dawson, C. W. Rowley, T. Colonius, B. J. McKeon, O. T. Schmidt, S. Gordeyev, V. Theofilis, and L. S. Ukeiley. Modal analysis of fluid flows: An overview. *AIAA J.*, 55(12):4013–4041, 2017.
- [35] D. H. Wolpert and W. G. Macready. No free lunch theorems for optimization. *IEEE T. Evol. Comput.*, 1(1):67–82, 1997.
- [36] Kai Fukami, Susumu Goto, and Kunihiro Taira. Data-driven nonlinear turbulent flow scaling with Buckingham Pi variables. *J. Fluid Mech.*, 984:R4, 2024.
- [37] Luca Guastoni, Alejandro Göemes, Andrea Ianaro, Stefano Discetti, Philipp Schlatter, Hossein Azizpour, and Ricardo Vinuesa. Convolutional-network models to predict wall-bounded turbulence from wall quantities. *J. Fluid Mech.*, 928:A27, 2021.
- [38] Zhicheng Wang, Dixia Fan, Xiaomo Jiang, Michael S. Triantafyllou, and George Em Karniadakis. Deep reinforcement transfer learning of active control for bluff body flows at high Reynolds number. *J. Fluid Mech.*, 973:A32, 2023.
- [39] Yifei Guan, Ashesh Chattopadhyay, Adam Subel, and Pedram Hassanzadeh. Stable a posteriori LES of 2D turbulence using convolutional neural networks: Backscattering analysis and generalization to higher Re via transfer learning. *J. Comput. Phys.*, 458:111090, 2022.
- [40] R. Bommasani, D. A. Hudson, E. Adeli, R. Altman, S. Arora, S. von Arx, M. S. Bernstein, J. Bohg, A. Bosselut, E. Brunskill, et al. On the opportunities and risks of foundation models. A report by the Center for Research on Foundation Models at the Stanford Institute for Human-Centered Artificial Intelligence, 2021.
- [41] Youngsoo Choi, Siu Wun Cheung, Youngkyu Kim, Ping-Hsuan Tsai, Alejandro N. Diaz, Ivan Zanardi, Seung Whan Chung, Dylan Matthew Copeland, Coleman Kendrick, William Anderson, Traian Iliescu, and Matthias Heinkenschloss. Defining foundation models for computational science: A call for clarity and rigor. arXiv:2505.22904, 2025.
- [42] J. N. Kutz and S. L. Brunton. Parsimony as the ultimate regularizer for physics-informed machine learning. *Nonlinear Dyn.*, 107(3):1801–1817, 2022.
- [43] S. L. Brunton and J. N. Kutz. Promising directions of machine learning for partial differential equations. *Nat. Comput. Sci.*, 4(7):483–494, 2024.
- [44] B. R. Rich. *Clarence Leonard (Kelly) Johnson 1910–1990: A Biographical Memoir*. National Academies Press, 1995.
- [45] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proc. Natl. Acad. Sci.*, 113(15):3932–3937, 2016.
- [46] Samuel H. Rudy, Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. Data-driven discovery of partial differential equations. *Sci. Adv.*, 3(4):e1602614, 2017.
- [47] Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In Sorelle A. Friedler and Christo Wilson, editors, *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, volume 81 of *Proceedings of Machine Learning Research*, pages 77–91. PMLR, 23–24 Feb 2018.
- [48] Bradley Efron and Trevor Hastie. *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*. Institute of Mathematical Statistics Monographs. Cambridge University Press, 2016.
- [49] Y. Li, E. Perlman, M. Wan, Y. Yang, C. Meneveau, R. Burns, S. Chen, A. Szalay, and G. Eyink. A public turbulence database cluster and applications to study Lagrangian evolution of velocity increments in turbulence. *J. Turb.*, 9(31):1–29, 2008.
- [50] E. Perlman, R. Burns, Y. Li, and C. Meneveau. Data exploration of turbulence simulations using a database cluster. Supercomputing SC07, ACM, IEEE, 2007.
- [51] Aaron Towne, Scott T. M. Dawson, Guillaume A. Brès, Adrián Lozano-Durán, Theresa Saxton-Fox, Aadhy Parthasarathy, Anya R. Jones, Hulya Biler, Chi-An Yeh, Het D. Patel, and Kunihiro Taira. A database for reduced-complexity modeling of fluid flows. *AIAA J.*, 61(7):2867–2892, 2023.

- [52] Wai Tong Chung, Bassem Akoush, Pushan Sharma, Alex Tamkin, Ki Sung Jung, Jacqueline H. Chen, Jack Guo, Davy Brouzet, Mohsen Talei, Bruno Savard, Alexei Y. Poludnenko, and Matthias Ihme. Turbulence in focus: Benchmarking scaling behavior of 3D volumetric super-resolution with BLASTNet 2.0 data. *Advances in Neural Information Processing Systems (NeurIPS)*, 36, 2023.
- [53] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T. Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J. G. Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A. C ’t Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons. The FAIR guiding principles for scientific data management and stewardship. *Sci. Data*, 3(1):160018, 2016.
- [54] Vishal Anatharaman, Jason Feldkamp, Kai Fukami, and Kunihiro Taira. Image and video compression of fluid flow data. *Theor. Comput. Fluid Dyn.*, 37(1):61–82, 2023.
- [55] Transparency (in training data) is what we want. *Nat. Mach. Intell.*, 7(3):329–329, 2025.
- [56] *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations (AIAA G-077-1998(2002))*. AIAA, 1998.
- [57] Cameron Tropea, Alexander L. Yarin, and John F. Foss. *Springer handbook of experimental fluid mechanics*. Springer, Berlin, 2007.
- [58] R. C. Smith. *Uncertainty quantification: theory, implementation, and applications*. SIAM, 2014.
- [59] Nick McGreivy and Ammar Hakim. Weak baselines and reporting biases lead to overoptimism in machine learning for fluid-related partial differential equations. *Nat. Mach. Intell.*, 6(10):1256–1269, 2024.
- [60] Steven L. Brunton, J. N. Kutz, Krithika Manohar, Aleksandr Y. Aravkin, Kristi Morgansen, Jennifer Klemisch, Nicholas Goebel, James Buttrick, Jeffrey Poskin, Adriana W. Blom-Schieber, Thomas Hogan, and Darren McDonald. Data-driven aerospace engineering: Reframing the industry with machine learning. *AIAA J.*, 59(8):2820–2847, 2021.
- [61] J. Tran, K. Fukami, K. Inada, D. Umehara, Y. Ono, K. Ogawa, and K. Taira. Aerodynamics-guided machine learning for design optimization of electric vehicles. *Commun. Engng*, 3:174, 2024.
- [62] Kai Fukami and Kunihiro Taira. Single-snapshot machine learning for super-resolution of turbulence. *J. Fluid Mech.*, 1001:A32, 2024.
- [63] Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed H. Chi, Quoc V. Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models. In *Proceedings of the 36th International Conference on Neural Information Processing Systems, NIPS ’22*, 2022.
- [64] Cristian Bodnar, Wessel P. Bruinsma, Ana Lucic, Megan Stanley, Anna Allen, Johannes Brandstetter, Patrick Garvan, Maik Riechert, Jonathan A. Weyn, Haiyu Dong, Jayesh K. Gupta, Kit Thambiratnam, Alexander T. Archibald, Chun-Chieh Wu, Elizabeth Heider, Max Welling, Richard E. Turner, and Paris Perdikaris. A foundation model for the earth system. *Nature*, 641(8065):1180–1187, 2025.
- [65] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017.
- [66] Dieter Jaksch, Peyman Givi, Andrew J. Daley, and Thomas Rung. Variational quantum algorithms for computational fluid dynamics. *AIAA J.*, 61(5):1885–1894, 2023.
- [67] Michael P. Brenner and Petros Koumoutsakos. Editorial: Machine learning and Physical Review Fluids: An editorial perspective. *Phys. Rev. Fluids*, 6:070001, Jul 2021.
- [68] Beverley McKeon, Eric Lauga, Brad Rubin, Arezoo Ardekani, Jackie Chen, Karthik Duraisamy, Sanjiva Lele, Julia Ling, George Rigas, and Peter Schmid. Editorial: Announcing Methods: New Experiments, Algorithms, and Theory (NEAT). *Phys. Rev. Fluids*, 9:110001, Nov 2024.