

# Harnessing Machine Learning to Guide Scientific Understanding

A clever use of machine learning guides researchers to a missing term that's needed to accurately describe the dynamics of a complex fluid system.

By **Sam Dillavou**

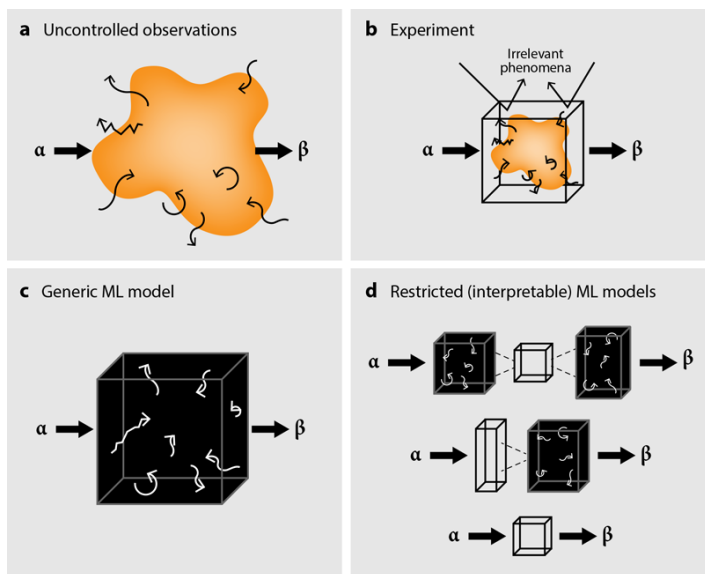
Physical theories and machine-learning (ML) models are both judged on their ability to predict results in unseen scenarios. However, the bar for the former is much higher. To become accepted knowledge, a theory must conform to known physical laws and—crucially—be interpretable. An interpretable theory is capable of explaining why phenomena occur rather than simply predicting their form. Having such an interpretation can inform the scope of a new theory, allowing it to be applied in new contexts, while also connecting it to and incorporating prior knowledge. To date, researchers have largely struggled to get ML models (or any automated optimization process) to produce new theories that meet these standards. Jonathan Colen and Vincenzo Vitelli of the University of Chicago and their colleagues now show success at harnessing ML not as a stand-in for a researcher but rather as a guide to aid building a model of a complex system [1]. In a demonstration of their method, the researchers have identified a previously overlooked term that leads to a more complete understanding of dynamics in a fluidic system.

To build new models, physicists often observe phenomena (Fig. 1a) in a controlled experiment (Fig. 1b) and attempt to relate parameters of the system to each other with equations. Then, through a combination of intuition and trial and error, they modify the experiment, the theory, or both until they find a set of equations that describes the data. Prior knowledge—for instance, that the system should have no history dependence, that temperature is uniform, or that gravity can be ignored—vastly shrinks the space of possible solutions and of required experimental exploration. This severe narrowing of scope is usually necessary for us humans, as we find it extremely difficult to grapple with a problem in more than a

handful of dimensions.

In contrast, ML models find more accurate and more generalizable solutions when given a (very) high-dimensional space to explore [2]. These models optimize enormous numbers of adjustable parameters until their predictions match the data. Unfortunately, the solutions found by generic ML models are often far too complicated and method dependent to extract a “why” [3]. Researchers applying such methods are therefore often limited to the unsatisfying claim that their data contains predictive information [4]. But what that information is and why it is predictive remains hidden in a black box of many messy variables (Fig. 1c). Techniques to identify where in the data that predictive information resides are emerging [5], but they are rarely used in the scientific process. An alternative to complex ML models is to use algorithms that search libraries of possible equations to describe a system directly [6]. However, this tactic scales poorly with system complexity, making it difficult to use on phenomena of modern interest. To incorporate ML into the general discovery process requires a balance: The method should have sufficient free rein to unlock its potential but also a restricted terrain on which the results will be interpretable.

Colen, Vitelli, and their colleagues now do just that using a sequence of ML algorithms [1]. Their work focuses on a paradigmatic problem in hydrodynamics: a single-file queue of water droplets in a microfluidic channel, suspended in a second fluid that causes them to interact and form a propagating shock front. This system has been previously modeled by a partial differential equation that describes the changing fluid density. But the equation, called Burgers' equation, fails to capture key



**Figure 1:** There are many approaches for understanding a physical phenomenon, where cause  $\alpha$  creates an effect  $\beta$ . (a) Uncontrolled observations prompt a loose connection between cause and effect. (b) A well-designed experiment combined with known physical laws leads to a simple cause–effect relationship, but designing such an experiment for complex phenomena can be extremely challenging. (c) A generic machine-learning (ML) model (neural network) optimizes many parameters to fit even complex data, but discerning the meaning of this black box of (for example, millions of) parameters is notoriously difficult. (d) ML models that include some restrictions or simpler operations (white) make it possible to identify certain aspects of the causal chain of complex phenomena to guide model building or further experiments.

Credit: S. Dillavou/University of Pennsylvania; APS/Alan Stonebraker

aspects of the system’s dynamics. To uncover the missing physics, the researchers first train an ML model to predict the time evolution of the 1D droplet density field  $\rho$ —in other words, they task their algorithm to find a function  $M$  that maps the initial density  $\rho_0$  forward in time:  $M[t, \rho_0] = \rho(t)$ .

To make their model interpretable, the researchers construct it from three successive operations. First, a neural network  $N$  transforms the density into a new 1D field, which they call  $\phi_0 = N[\rho_0]$ . While this “latent” field does not have an easily interpretable physical meaning, it only contains information about the initial density field. Second, this field is fed into a

function, called  $F$ , that steps it forward in time—in other words,  $F(\phi_0, t) = \phi(t)$ . The researchers restrict the form of  $F$  to a set of linear operations. Finally, the field is transformed back into density by another neural network, essentially an inverse of the first step. (Mathematically, the entire process can be described as  $M[t, \rho_0] = N^{-1}[F(t, N[\rho_0])] = \rho(t)$  and is drawn schematically in Fig. 1d, top). By simultaneously optimizing all three steps to match experimental data, the researchers found better predictions than those made by Burgers’ equation.

The researchers then utilized an algorithm that finds simplified analytical approximations of numerical functions [6]. This step would fail for a typical neural network trained on the experimental data (Fig. 1c). But, notably, it produces a five-term linear partial differential equation as a good stand-in for  $F$ . Despite this equation operating on the (uninterpretable) latent variable  $\phi_0$ ,  $F$ ’s role as a time propagator makes each term’s meaning intelligible at a high level. Specifically, the researchers identify one of the differential terms as connected to dispersion—a frequency dependence in the wave speed of the fluid. Such a dispersive term is not present in Burgers’ equation, but the team found that its addition produces more accurate descriptions of the shock-front dynamics that arise in the droplet density field. Finally, the team develops a model of interacting droplets and finds that this added dispersive term is a direct consequence of nonreciprocal hydrodynamic interactions.

This work provides an exciting use of ML as a compass during scientific exploration, which requires a fundamentally different approach than standard ML practice, where models are judged primarily by their prediction accuracy. However, for scientific exploration the “best” models are the ones that lead to physical insight (the “why”) but may not be the most accurate. In fact, the team found that adding the key dispersive term actually raised the predictive error slightly compared to other ML models applied to the same problem; however, it clearly captured missing physics occurring near the shock front. Rather than lower error, it was closing the loop with a continuum model and identifying the source of this dispersive term that allowed Colen, Vitelli, and colleagues to solidify their conclusions. This workflow dovetails with recent work here at the University of Pennsylvania using ML as an experimental guide [7], wherein the simplest and “weakest” (least-predictive) models trained to predict clogging in granular materials gave the most insight,

prompting experiments that solidified their interpretation.

Increases in computational power have massively accelerated analysis of scientific data, yet our exploration of that data often remains entirely human driven. As physicists study increasingly complex emergent phenomena, the dimensions of potential physical models and therefore the complexity of required experimental exploration grows rapidly. While standard analysis tools allow us to identify robust trends, it may not be feasible to hunt down highly nonlinear, history-dependent, and multiscale effects in (necessarily) messy data without a guide capable of ingesting 100 dimensions at once. To study such phenomena, fluency with both the subject matter and ML tools may prove an invaluable combination, both as an experimental guide and a theoretical one.

**Sam Dillavou:** Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, PA, US

## REFERENCES

1. J. Colen *et al.*, “Interpreting neural operators: How nonlinear waves propagate in nonreciprocal solids,” *Phys. Rev. Lett.* **133**, 107301 (2024).
2. J. W. Rocks and P. Mehta, “Memorizing without overfitting: Bias, variance, and interpolation in overparameterized models,” *Phys. Rev. Res.* **4**, 013201 (2022).
3. C. Rudin *et al.*, “Interpretable machine learning: Fundamental principles and 10 grand challenges,” *Statist. Surv.* **16**, 1 (2022).
4. S. Dillavou *et al.*, “Beyond quality and quantity: Spatial distribution of contact encodes frictional strength,” *Phys. Rev. E* **106**, 033001 (2022).
5. K. A. Murphy and D. S. Bassett, “Information decomposition in complex systems via machine learning,” *Proc. Natl. Acad. Sci. U.S.A.* **121**, 13 (2024).
6. S. L. Brunton *et al.*, “Discovering governing equations from data by sparse identification of nonlinear dynamical systems,” *Proc. Natl. Acad. Sci. U.S.A.* **113**, 3932 (2016).
7. J. M. Hanlan *et al.*, “Cornerstones are the key stones: Using interpretable machine learning to probe the clogging process in 2D granular hoppers,” [arXiv:2407.05491](https://arxiv.org/abs/2407.05491).