



Pervasive machine learning in physics

No longer restricted to data analysis, machine learning is now increasingly being used in theory, experiment and simulation — a sign that data-intensive science is starting to encompass all traditional aspects of research.

In a talk given in January 2007, computer scientist Jim Gray, recipient of the Turing Award, outlined the paradigms of science: first, empirical observations, second theory, third, computation. He coined the idea of a fourth paradigm: discovery enabled by the exploration of massive datasets¹. “The techniques and technologies for data-intensive science are so different that it is worth distinguishing data-intensive science from computational science as a new, fourth paradigm for scientific exploration.” Gray suggested that the fourth paradigm unifies theory, experiment and simulation — he was right in a way he unfortunately did not get the chance to witness: later that month he was lost at sea.

Whether one chooses to trace back the links between machine learning and physics 40 years ago to John Hopfield and his Ising model of a neural network, or rather focus on more recent statistical mechanics insights into deep learning², connections between the two fields run deep. Physicists have been early users of machine learning methods in data analysis, well before the advent of deep learning around 2012. For example, machine learning was already [discussed](#) at meetings in high-energy and nuclear physics in 1990, with an earlier suggestion for the potential use of neural networks in experimental particle physics³.

The first, obvious use cases were in the analysis of the growing volumes of experimental data, a topic of ongoing focus (see this [Review](#) in this issue). As the technology advanced and the accessibility and ease of use of machine learning tools improved, other aspects of research have started to rely on them: experiment design and the optimization of operating parameters, data acquisition and pre-processing. These applications are related to data analysis in the sense that they need to efficiently explore high-dimensional parameter spaces, where human intuition is not necessarily a good guide.

What is perhaps less obviously related to data analysis is the use of machine learning in simulation, numerical computation and theory. For example, physics-informed neural networks (see this [Review](#)) are driving new advances in fluid dynamics simulations, a field that has also been exploring the uses of machine learning since the 1990s. In a [Comment](#) in this issue, Ryan Pederson and colleagues overview the advances in using machine learning in density functional theory and ask whether

there is still a place for human insight and intuition in designing functionals. In a [Comment](#) about the use of machine learning in mathematics and theoretical physics, Michael R. Douglas suggests that the combination of recent developments can be “as revolutionary for science as was the original development of scientific computation and simulation,” a view echoing Gray’s definition of the fourth paradigm.

With improved capability also come challenges. Examples include the integration of machine learning throughout experimental processes as outlined in this [Comment](#); interpretability, discussed in this [Comment](#); and the choice of methods given the wide availability of algorithms, addressed by the introduction of scientific benchmarks discussed in a [Perspective](#) in this issue.

We have collected Reviews, Comments and Tools of the Trade articles, illustrating the wide spectrum of machine learning applications in physics and discussing trends and challenges. We hope that this ongoing [series](#) will become both a useful resource and the start of a dialogue among theorists, experimentalists and computational scientists from different areas of physics. We also hope to offer in our pages a forum for discussions on topics such as integration, verification and validation, benchmarking, interpretability or uncertainty estimation.

Gray did not witness the full impact demonstrated by machine learning in science, yet he was absolutely right about theory, experiment and simulation being unified through data-intensive science, albeit perhaps in a different way than he imagined 15 years ago. There is no reason why the fourth paradigm should be the last. As technology and human knowledge expand and evolve so will the ways we do science. A natural question is what will the fifth paradigm look like (and when will it arrive). Perhaps we already can foresee it: machines being no longer mere tools, but equal partners in scientific exploration, exchanging ideas, intuition and understanding with the human peers. This vision may be closer than we imagine.

1. Hey, T., Tansley, S. & Tolle, K. *The Fourth Paradigm: Data-Intensive Scientific Discovery* (Microsoft Research, 2009).
2. Bahri, Y. et al. Statistical mechanics of deep learning. *Annu. Rev. Condens. Matter Phys.* **11**, 501–528 (2020).
3. Denby, B. Neural networks and cellular automata in experimental high energy physics. *Comput. Phys. Commun.* **49**, 429–448 (1988).