

# Physics Is Different: Context, Culture & Craft in Effective AI for Physics

*Tess Smidt*

*Physics is a sprawling and ambitious enterprise unified by the goal of explaining the natural world, yet it is extraordinarily diverse in how that goal is pursued. The opportunities to harness artificial intelligence to aid in our pursuits to understand the universe are vast. But the heterogeneity of physics in practice is a vivid demonstration of how scientific inquiry varies across scale, infrastructure, and standards of evidence, and how machine learning integrates unevenly as a result. While current applications of AI often focus on accelerating existing workflows, the deeper promise lies in developing robust methods that plug into physicists' established workflows and toolkits; lower barriers in the continuous cycle between measurement, simulation, analysis, and theory; and create fundamentally new ways of probing, modeling, and controlling physical systems.*

**P**hysics is big. Really big. It is both the science of first principles and the practice of probing every corner of the natural world, from subatomic particles to the cosmos. The word itself suggests unity, but in practice the field is anything but monolithic. Particle physicists build colliders the size of cities to glimpse fleeting quarks and gluons. Condensed matter physicists coax exotic behaviors out of crystals grown milligram by milligram. Astrophysicists study a universe that can never be rerun. Plasma physicists wrestle with the controlled chaos of hot, charged matter. What unites these efforts is an ambition to explain how the world works, and a shared toolbox of habits – reductionism, emergence, approximation, symmetry, unlearning – that make the complexity of nature tractable.

How artificial intelligence enters these pursuits is bespoke and uneven. Artificial intelligence techniques already accelerate simulations, parse noisy detector streams, and help steer experiments in real time. In some subfields, they are embedded deeply into workflows; in others, their role remains speculative. There is a popular temptation to treat AI as a universal solvent, an inevitable engine of automated discovery. But physics offers a cautionary tale. The field's extraordinary diversity of data, infrastructure, and culture means there is no single interface for machine

learning to plug into. The tools that succeed will amplify human judgment, stay grounded in physical context, and enrich the shared work of interpretation, helping us frame new questions and broaden what counts as inquiry.

This essay is an attempt to map that terrain. I begin by sketching the varied enterprise of physics itself, highlighting the activities that define it – experiment design, measurement, simulation, analysis, theory, and modeling – and the infrastructures that sustain those activities. I then turn to concrete vignettes, from the fragile ripples of gravitational waves to the torrents of data at the Large Hadron Collider (LHC) to the patchwork of experiments and simulations in atomistic science, to show how machine learning is already entangled with physics in practice. From there, I consider the toolboxes of physics and of machine learning and the resonances that have begun to emerge between them. Finally, I explore the challenges and frontiers ahead: what it will take for AI not merely to speed up existing pipelines, but to help us articulate new abstractions, reframe problems, and expand what it means to understand.

The message I hope to leave you with is this: physics will not be “solved” by AI. But AI can become an essential part of the physicist’s toolkit, lowering barriers between experiment design, measurement, simulation, and modeling, and giving us new knobs to probe and play with the universe.

**P**hysics has a reputation for elegance, compact equations like  $E = mc^2$  or  $F = ma$  that seem to distill reality into symbols. But behind those icons lies an extraordinary range of practices. Doing physics means designing experiments and instruments, making measurements, running simulations, analyzing data, and modeling phenomena. No single physicist traverses all of these steps; the field is sustained by collaborations in which each specialist contributes deep expertise to a narrow part of the cycle, with just enough overlap to communicate across boundaries.

These activities are worth outlining not only to show the variety of what physicists do but also to set the stage for how artificial intelligence might help at each step.

*Design.* Often, physicists must first create the very systems they hope to study. This can mean assembling lattices of ultracold atoms on an optical table or growing crystals grain by grain in a lab furnace. It also includes design of the experimental tools needed to capture signals, such as fabricating detectors that hold thousands of tons of liquid argon to entice elusive neutrino interactions or kilometers of arrays of antennas looking at the spectra of the cosmos.<sup>1</sup> Experiment design pushes the limits of engineering to expand what questions can even be asked; what we can build shapes the scope of inquiry long before any measurement is made.

*Measurement.* Measurement captures signals from the world, whether naturally occurring or carefully contrived. At one extreme are kilometer-scale interferometers like the Laser Interferometer Gravitational-Wave Observatory (LIGO) that

detect spacetime ripples a thousand times smaller than a proton's width.<sup>2</sup> At the other extreme are tabletop spectrometers that fit on an optical bench and reveal atomic transitions through the absorption or emission of light. To a physicist, there is no hard line between system and instrument: every observation is an interaction between physical actors. Even the simplest measurements are mediated by probes (made of photons, electrons, neutrons, and so on) that shape the result.

*Simulation.* When direct measurement is impossible, physicists turn to simulation: structured guesswork that explores what might happen under given laws. Early Monte Carlo methods, invented at Los Alamos, were designed to track neutrons through fission reactions.<sup>3</sup> Today, simulations extend across scales: particle physics detectors modeled in exquisite detail; universes evolved under different cosmological parameters; atomic trajectories stepping through femtoseconds of molecular motion. Simulations are not "truth machines" but approximations, balancing fidelity against computational cost. The assumptions built into simulations – what physics to include, what variables to simplify – matter as much as the simulation's outputs.

*Analysis.* Raw signals or simulated traces rarely speak for themselves. Analysis transforms them into evidence through statistics, signal processing, calibration, and inference. Consider the Large Hadron Collider: thirty million proton-proton collisions per second, most discarded within microseconds by custom electronics and fast classifiers.<sup>4</sup> Or cosmological surveys, in which galaxies must be disentangled from atmospheric noise, detector artifacts, and foreground contamination. Analysis is where judgment enters most visibly, deciding which patterns are significant, which uncertainties dominate, and which anomalies to trust.

*Theory and modeling.* Physics is perhaps best known as a science of first principles. Its iconic laws – Maxwell's equations, quantum mechanics, general relativity, the Standard Model – suggest a discipline governed by elegant, universal rules. These are what physicists call theories: compact statements about how the world should behave, often formulated with remarkable mathematical economy.

Yet those pristine equations are rarely sufficient on their own. They describe everything, and therefore almost nothing in particular. Turning theory into a working tool requires modeling, constructing simplified or approximate systems that make the consequences of theory concrete. Modeling translates universal law into a form that can be calculated, compared, and understood. The Ising model distills the essence of phase transitions; Ginzburg-Landau theory captures superconductivity through a coarse-grained order parameter; simplified cosmologies let us test how inflation or dark matter might leave observable traces. A good model makes the overwhelming interpretable, showing which interactions or symmetries truly matter.

Reductionism and emergence guide this process from opposite ends. Modeling often begins by reducing complex phenomena into simpler parts, yet the rich-

est insights appear when new collective behaviors emerge that the fundamental equations never single out. Symmetry sits between them. It describes invariance, the aspects of a system that remain unchanged under transformation; and when symmetry is broken, it reveals how new structure or order can arise. Principles like these help physicists reason about systems through constraints and conservation laws rather than by tracking every microscopic detail.

Approximation is a constant companion for understanding a complex world. Almost no real-world problem can be solved exactly. Physicists linearize equations, expand functions in series, or integrate out degrees of freedom to make problems tractable. In many cases, knowing which details to ignore is as important as knowing which to keep. Yesterday's laws often survive as limiting cases of broader ones: Newtonian mechanics within relativity, classical electrodynamics within quantum field theory.

Theory provides the general rules; modeling provides the playground where those rules become intelligible. Machine learning faces a similar challenge. Neural networks can, in principle, represent extraordinary complexity, but to make sense of what they learn requires added structure and interpretation. This should not be surprising; even in physics, compact laws alone do not yield intuition. What turns theory into insight is the scaffolding that manages complexity and invites play: the approximations, symmetries, and models that let us explore how systems behave when their rules are bent or simplified.

A good scientific tool, in this sense, is not one that simply confirms what we expect but one that has the authority to surprise us. The same must be true for new tools of artificial intelligence. To contribute meaningfully, they cannot bypass the habits that make understanding possible; they must work within them or help us find more powerful paradigms. The most promising uses of AI in physics will be to build tools that amplify and reorganize the familiar cycles of design, measurement, simulation, analysis, theory, and modeling.

*Interdependence and infrastructure.* Though described separately, these activities (design, measurement, simulation, analysis, theory, and modeling) are tightly intertwined. Simulations depend on models; analyses depend on simulations; measurements depend on design; theories are revised by all. And none of this happens in a vacuum. Infrastructure – whether a shared beamline, a national supercomputer, or a tabletop laser – sets both the feasibility of questions and the barriers to asking them. In particle physics, data volumes are so immense that experiments resemble institutions more than instruments. In condensed matter, the bottleneck may be access to a single synchrotron beamline for a few hours of precious time.

**G**ravitational waves are ripples in spacetime produced when massive objects accelerate asymmetrically, for example, black holes or neutron stars orbiting each other in an ellipse. Einstein predicted gravitational waves in 1916;

they were first observed a century later, in 2015, by LIGO.<sup>5</sup> Detecting them requires astonishing precision, measuring changes in distance a thousand times smaller than the width of a proton.

Each of the two LIGO sites consists of a giant four-kilometer-long interferometer that splits a laser beam into two perpendicular arms. A passing gravitational wave stretches one arm while squeezing the other, altering the distance the light travels. For a measurement to count, both sites must be able to measure the same ripple. The challenge lies in the faintness of the signal and the overwhelming background of noise. The detectors are exquisitely sensitive to any vibration, whether it be earthquakes, air conditioners, thermal drift, lightning strikes, even cosmic rays.<sup>6</sup> Keeping the mirrors aligned and the instrument stable is key for a meaningful measurement.

Experimental time is too valuable to risk on untested methods, so most are proven offline to earn a role in live operations. And because measuring something unprecedented demands precision and interpretability, AI might seem an unlikely tool. But when carefully targeted, machine learning has already proven useful.

For instance, machine learning has been used to improve mirror control, suppressing vibrations and technical noise, and to stabilize the light injection system, particularly the delicate quantum “squeezing” process that reduces photon noise at high frequencies.<sup>7</sup> These approaches show how AI can enhance critical control systems, extending the detectors’ reach.

Machine learning is also now part of LIGO’s real-time detection pipeline. For example, the Aframe system uses deep neural networks trained on millions of simulated waveforms mixed with real detector noise to identify gravitational-wave signals within seconds of their arrival.<sup>8</sup> Its performance rivals traditional matched-filter searches while running far faster and with lower computational cost. The companion system AMPLFI applies similar methods to estimate the properties of each event, such as the masses and location of merging black holes, almost immediately.<sup>9</sup> Together, these AI-based pipelines now issue public gravitational-wave alerts in real time, as for the recent event S250830m, very likely a binary black hole merger.<sup>10</sup>

A team of researchers from the Max Planck Institute for the Science of Light and the LIGO Laboratory at Caltech used AI not to run detectors but to reimagine them.<sup>11</sup> By searching the immense design space of possible interferometer layouts with a “universal interferometer” model and large-scale physics-based simulations, they discovered dozens of novel topologies that outperform current next-generation designs in sensitivity. While these ideas face major hurdles of cost, engineering, and implementation, they demonstrate how AI can propose new directions for the future of gravitational-wave detection.

A common thread of all these applications of machine learning for gravitational-wave detection is that it is bespoke. While elements might step from other applica-

tions, the end-to-end algorithm (the input data, architecture, loss functions, and training scheme) needed to be built from scratch with careful consideration of the task at hand. In that sense, gravitational-wave detection epitomizes both the promise and the challenge of AI in physics: a tool that can enhance sensitivity and accelerate discovery, but only when deeply entangled with the infrastructure and ethos of the field.

**W**hen the objects of study are fundamental particles – the quarks, leptons, force mediating bosons from the Standard Model or beyond – the only option is to watch them interact. Particle colliders help do this by smashing particles together at near-light speeds:  $E = mc^2$  in action. The faster and more energetic the collisions, the more new particles are created from the collision.

Particle physics experiments, particularly those at colliders, generate some of the largest and most homogeneous datasets in science. The homogeneity comes from control: every proton-proton collision occurs under nearly identical conditions, unlike the irreproducible complexity of most natural systems. But with such enormous volumes of data, even rare statistical fluctuations can start to look meaningful. An observation at  $3\sigma$  significance, 99.7 percent confidence – which is as unlikely as flipping a coin and getting all heads in eleven tosses – will appear somewhere in billions of collisions simply by chance. To guard against mistaking these coincidences for discoveries, physicists set the bar at  $5\sigma$  significance, 99.99994 percent confidence, before declaring a new particle or phenomenon to be real.

At the Large Hadron Collider, detectors see up to thirty million collisions per second. Storing everything is impossible, so experiments rely on a multilayered pipeline to select the most informative events. A hardware trigger, built from custom electronics including field-programmable gate arrays (FPGAs), selects about one in four thousand collisions within microseconds. Surviving events pass to software triggers running on vast CPU and GPU farms, which apply more detailed criteria. Only then are the remaining candidates reconstructed: raw detector hits assembled into tracks, tracks combined into jets, and objects tagged as electrons, photons, or heavy-flavor decays. Continuous calibration and alignment account for detector drift, while anomaly detection safeguards against both faulty signals and unexpected physics. To hedge against hidden biases, a small random subsample of untriggered events is also stored.

Machine learning has supported this pipeline for decades. Neural networks were already used for pattern recognition in the 1990s, long before “deep learning” became fashionable.<sup>12</sup> What has changed is the scale and specialization of the models. Today, ultrafast classifiers run directly on FPGAs, graph neural networks find tracks in dense collision environments, convolutional networks suppress pile-up noise, and autoencoders assist in calibration and monitoring.<sup>13</sup> Many of these models are trained on simulated detector responses using tools such as GEANT4 that

provide the labeled data necessary for supervised learning. Because this introduces the risk of mismodeling, collaborations continually validate against control samples, retrain on real collisions, and parameterize known sources of uncertainty.<sup>14</sup>

Machine learning now underpins nearly every layer of the data pipeline, operating under some of the most stringent real-time requirements in science. In the first-tier trigger, custom models implemented on FPGAs or application-specific integrated circuits execute with latencies below fifty nanoseconds, filtering millions of events per second; higher-level triggers and offline analyses then refine those selections through increasingly complex inference.<sup>15</sup> The result is one of the most elaborate and rigorously validated deployments of AI in any scientific domain.

Yet its success depends as much on culture as on computation. The LHC community prizes redundancy, interpretability, and continuous cross-checks, values that make it possible to trust automation at such scale. Far from replacing physics expertise, AI is absorbed into an infrastructure of verification honed over decades. The LHC epitomizes one archetype of physics data: immense, controlled, homogeneous, and sifted through industrial-scale analysis pipelines. Machine learning thrives here not by offering general-purpose models but by becoming a set of specialized tools tailored to the unforgiving statistical and cultural demands of particle physics.

**O**n a scale larger than gravitational waves and particle collisions but still invisible to the naked eye lie atoms and the systems they form: molecules, crystals, proteins, nanoclusters. These systems are the playground of condensed matter physics, materials science, and chemistry, fields that vary more in the types of questions they're interested in than in their subject matter. Atomic systems illustrate a different face of physics, one rich in methods and degrees of control but fraught with inconsistency and approximation.

Atoms can, in principle, be arranged, synthesized, and probed in myriad ways. The atomistic sciences are overflowing with an alphabet soup of characterization techniques: X-ray diffraction (XRD) for crystallography, transmission electron microscopy (TEM) for atomic imaging, angle-resolved photoemission (ARPES) for electronic structures, nuclear magnetic resonance (NMR) for local environments, and many more. Some of these tools fit on a lab bench, others demand infrastructure the size of a city block. Synchrotrons accelerate electrons around kilometer-scale rings to generate intense X-rays. Free-electron lasers are so bright they vaporize the very sample they probe. Neutron sources offer complementary information by scattering particles without charge. Even magnets (the strongest sometimes drawing 10 percent of a city's power) become instruments to test matter under extreme conditions.<sup>16</sup>

This diversity is a blessing and a curse. There are plenty of data, but they are patchy, inconsistent, and often specialized. No two probes see exactly the same

thing, and experimental fluency is required to navigate across techniques. This patchwork of formats, conventions, and partial coverage makes it extremely difficult to build cohesive benchmarks

But if experiments are diverse, the underlying theory seems, at first glance, wonderfully unified. The Schrödinger equation, in principle, describes all quantum systems of electrons and nuclei. But in practice it can only be solved exactly for hydrogen. Even helium requires approximation. For anything larger (molecules, crystals, surfaces), physicists descend a hierarchy of methods: quantum Monte Carlo, coupled cluster, Hartree-Fock, and density functional theory (DFT). Each method trades off accuracy, interpretability, and computational cost.

The workhorse is DFT, which balances tractability with reasonable fidelity. It has enabled vast databases of calculated properties and powered high-throughput searches for new materials.<sup>17</sup> Yet it is far from push-button. Every system demands choices of functionals, basis sets, and convergence criteria; no single setup works universally. The field relies on a diverse ecosystem of DFT codes. Ensuring they produce consistent results has required deliberate coordination: a major benchmarking effort a decade ago confirmed that, with care, independent implementations agree to within a few millielectronvolts per atom.<sup>18</sup>

Fortunately, DFT's approximations are largely systematic; standard methods underestimate band gaps and misrepresent some electron correlations. These biases become part of the interpretive culture of the field: researchers know which properties to trust and which to treat skeptically.

Machine learning enters atomistic science in two complementary ways. On the experimental side, AI is now used to interpret noisy spectra, accelerate image analysis in electron microscopy, and guide beamline experiments in real time.<sup>19</sup> At large facilities, the deluge of data – petabytes from experiments running at synchrotrons, neutron sources, or free-electron lasers – makes automated analysis not just convenient but essential.<sup>20</sup>

On the simulation side, neural networks trained on quantum mechanical calculations act as surrogates. “Machine-learned interatomic potentials” can approximate energies and forces with near-DFT accuracy but orders of magnitude faster, enabling molecular dynamics of systems thousands of times larger or longer-lived than conventional calculations allow. Graph and Euclidean neural nets embed physical symmetries directly into their architectures, ensuring that predictions of atomistic systems respect invariance under permutation or rotation, respectively.<sup>21</sup> These constraints are not decorative: without them, the models are less reliable and often less efficient. Because symmetries pervade physics, equivariant models have also found use in other domains, including high-energy physics.<sup>22</sup>

Notably, the training data for these surrogates are themselves shaped by approximation. A model built on DFT inherits DFT's biases. It may be fast and accurate on its terms, but is only as trustworthy as the assumptions encoded in its source.

This is why validation against experiments, or cross-checks with higher-level theory, remains indispensable.

Atomistic systems thus present a paradox. They offer tremendous experimental freedom, myriad ways to probe matter, and a unifying theoretical foundation in quantum mechanics. But in practice, the data are messy and the theory is approximated, leaving wide gaps for interpretation. Machine learning is valuable precisely because it can bridge across these mismatches: extending the reach and fidelity of simulations, denoising spectra, interpolating between scarce experiments. At the same time, it inherits the fragility of its inputs. In atomistic science, AI is not a magic key but a tool that thrives only when tethered tightly to physical insight and cross-validation.

**T**he path forward is not about replacing the physicist's craft but making it easier to use and extend. The real opportunity lies in creating systems that let ideas, data, and methods travel more freely – between instruments and analysis pipelines, between subfields that rarely share language, and between the computational and experimental parts of the discipline. This need for connective infrastructure and shared abstractions echoes the conclusions of the recent NSF Workshop on the Future of Artificial Intelligence and the Mathematical and Physical Sciences, which emphasize integration across the full scientific cycle – from data acquisition to theory – rather than isolated algorithmic advances.<sup>23</sup>

Progress often depends on small, local acts of translation: building the glue code that connects detectors to data streams, theory to simulation, or one community's conventions to another's. The next step is to make that translation easier and more durable. That means developing abstractions that encode physical intuition without demanding heroics, infrastructures that let existing setups plug into modern data workflows, and frameworks that make the growing body of scientific knowledge more searchable and interoperable. Only once those foundations are in place can we build models capable of higher abstraction, models that reason across scales and modalities, whose reliability rests on the very interconnectedness of the system. The challenge and the promise are the same: to make physics more continuous, cumulative, and playful by building AI systems that extend how we think and experiment.

**A**t the moment, every functional application of AI in physics rests on deep, local expertise. These models succeed because physicists and computer scientists have embedded physical structure directly into them: preserving symmetries, constraining transformations, encoding the logic of experiment, and validating until skepticism is satisfied. This is not automation; it is modeling by another name.

But there is a cost to customization. Performant models are rarely “plug and play.” In atomistic modeling, for example, productive use requires fluency in mul-

multiple domains at once: training and validating neural networks, generating and interpreting quantum calculations such as DFT, running molecular dynamics, and connecting atomic trajectories to experimental observables. That combination of skills is uncommon, and without safe abstractions, every new model must be built nearly from scratch. The DFT code VASP succeeded not by replacing expertise but by embedding it – decades of tuning functionals and convergence criteria – so researchers could build on reliable defaults without mastering every detail. Machine learning in physics still lacks such durable interfaces.

Furthermore, there is a friction between custom architectures with useful abstractions and computation performance. Equivariant neural networks that encode symmetry through tensor products or group representations are a striking case in point. These methods are mathematically demanding and map poorly onto GPU hardware optimized for dense matrix multiplications. Equivariant and physics-informed architectures often run slower and scale worse than their generic transformer cousins, not because they are flawed but because far less engineering effort has gone into making them efficient.

Bridging this gap falls to researchers and students who sit between disciplines. They tune architectures, reconcile units and conventions, and write the glue code that lets symbolic physics talk to numerical data. Each “AI breakthrough” in physics hides a scaffolding of bespoke pipelines and late-night debugging sessions. These efforts produce remarkable results, but they are fragile. If a few key people move on, the knowledge of how to make things work can vanish.

Fragmentation compounds the problem. Physics is not one field but many, each with its own data culture, instruments, and timescales. In some domains, like particle physics and cosmology, massive homogeneous datasets demand industrial-scale pipelines. In others, like condensed-matter physics or quantum information, data are heterogeneous and deeply contextual. Even defining a benchmark is hard: every experiment has its own geometry, calibration, and noise characteristics, and analysis tools evolve alongside the instruments themselves. The meaning of any dataset is inseparable from how it was made.

The near-term challenge is to make this integration less punishing, to design environments where physics and machine learning can meet without so much heroic labor. That will mean building abstractions that carry domain knowledge forward so future students can explore new ideas rather than constantly rebuilding the same ones. Ultimately, success will look less like end-to-end automation and more like the emergence of robust, trustworthy interfaces between experiment, simulation, and learning.

**B**uilding tools that observe the world is difficult, slow, and expensive. The design, coordination, and construction needed for particle accelerators, telescopes, synchrotron beamlines, and new generations of microscopes

often take decades. Their purpose is to reach new regimes of measurement, making signals visible that were previously inaccessible. Progress in physics depends as much (if not more) on these instruments as on the theories that interpret them.

Artificial intelligence, by contrast, evolves on the timescale of months. New architectures, training schemes, and software frameworks appear faster than any experimental cycle could ever adapt. This creates a widening gap between how quickly AI develops and how slowly physical infrastructure can evolve. Cryogenic electron microscopy (cryo-EM) is a vivid example. The method has been in development since the 1970s, but only after advances in detector hardware and reconstruction algorithms did it achieve near-atomic resolution, earning a Nobel Prize in 2017.<sup>24</sup> Now, just as the hardware has matured, the computational stack beneath it is being rewritten again: motion-correction, denoising, and 3D-reconstruction algorithms are rapidly incorporating machine learning. Cryo-EM shows how long hardware takes to stabilize and how quickly the software layers atop it can shift, forcing scientists to rebuild analysis pipelines even as the microscopes themselves remain unchanged.

This mismatch recurs across physics. Instruments that take decades to build are paired with software that reinvents itself every few months. For now, the most meaningful role of AI is therefore to help us make better use of the tools we already have: to coordinate experiments, interpret results, and connect observations into larger bodies of knowledge.

Even discovering what data exist can be arduous: datasets are scattered across facilities and formats, and their interpretive context often lives only in lab notebooks or custom scripts. Depositing data or creating benchmarks requires documenting calibration, noise, and uncertainty in exhausting detail. AI could ease some of that burden by learning to recognize overlap among experiments, automate the collation of data streams, and link partial results into a coherent map of what has been observed. A system that can suggest the next most informative experiment or identify patterns across instruments would make scientific progress cumulative rather than episodic. In this sense, the frontier is not just faster computation but better coordination, building connective tissue that allows insights to travel between experiments, institutions, and scales.

Even our accumulated knowledge faces similar coordination problems. The scientific literature contains centuries of data, models, and insight, yet much of it is effectively dark matter, too vast and unevenly indexed for any individual to navigate. Large language models already hint at how this might change. As scientific literature assistants, they could help researchers trace relationships across subfields, summarize prior work, and connect results that use different vocabularies or conventions. The technology exists now; what is missing is reliability and integration. Modern systems still invent citations, blur context, and struggle with paywalls and proprietary archives. Building assistants that respect provenance, verify sources,

and work within the publishing ecosystem could make the literature itself a searchable, dynamic experiment wherein ideas are as accessible as data.

Even more transformative would be using these tools to translate across disciplinary boundaries. Every subfield of physics and every neighboring science carries its own language, priorities, and sense of what counts as progress. A discovery in one area may be routine in another; a null result in one domain may be a signal in a different framing. Large models trained with and guided by social scientists and historians of science could make these distinctions legible, mapping not just concepts but the cultures that produced them. Reinventing the wheel is the tragedy of the unsearchable; AI, used thoughtfully, could help us see where the wheels already exist, why they were built the way they were, and what else they could be used to move.

Once AI systems help integrate context and coordinate knowledge, they can begin to expand what kinds of physical systems we can represent and reason about. The goal is not simply to make existing workflows more efficient, but to reach the gray zones between our simplifying regimes, places where theory, experiment, and simulation have each struggled to stand alone.

Multiscale modeling is one of these frontiers. Physical theories and computational tools are typically built for specific domains: quantum mechanics for electrons, molecular dynamics for atoms, continuum mechanics for materials. Between them lie regimes where the spatial and temporal separations that justify standard approximations break down, and where small-scale structure and collective behavior become inseparably intertwined. Machine learning can help bridge these seams, learning mappings between levels of description or inferring effective parameters directly from data. In doing so, it can connect first-principles calculations to emergent phenomena without the loss of interpretability that often comes from coarse-graining.

Another challenge is multimodality. The same system can be observed in many incompatible ways, through diffraction, microscopy, spectroscopy, or transport, each yielding a partial and often biased picture. AI could help integrate these perspectives, learning how one type of measurement constrains another or how multiple views combine into a consistent description. Such models could even propose new directions for investigation, identifying which probes or conditions might most reduce uncertainty or reveal an unmeasured degree of freedom.

These capabilities mark a qualitative shift: from AI as a post-hoc analyst of data to AI as an active participant in the modeling process. A multiscale, multimodal model does not replace theory or experiment; it acts as connective tissue between them. It can represent phenomena too rich for any single equation yet too structured to be left to blind fitting.

Generative models already play a role in guiding experiments, but only in narrow, well-structured settings in which automation is feasible and objectives are

clearly defined. Models can propose candidate molecules, crystal structures, or experimental parameters, and such tools have begun to accelerate discovery in condensed matter physics, chemistry, and materials science. Yet these systems still operate within fixed contexts: they explore what is already measurable rather than helping us recognize what could be measured – phenomena or configurations we have not yet thought to test.

AI that can plan or orchestrate experiments is a recurring scientific dream, one that always feels five or ten years away. The difficulty is not only conceptual but physical. True experimental orchestration requires laboratories, instruments, and workflows designed for continuous machine-human interaction or for no human interaction at all, except when the machine inevitably breaks, because scientific instruments are never as robust as we would like them to be. Changing those environments is slow and capital-intensive, far harder than deploying new software. What sounds like a modest step, letting models choose the next experiment, is in fact one of the most ambitious because it touches the very infrastructure of how we do science.

The deeper promise is for AI that engages with experiment and simulation as a reasoning partner, able to weigh uncertainty, recognize when a model's assumptions fail, and suggest new hypotheses rather than merely new samples. Such tools would help researchers explore vast design spaces and build intuition about them. As these systems mature, AI's role in physics will evolve from tool to collaborator. Models that can reason across scales and modalities will not just analyze data but participate in theory-building, connecting disparate observations through shared structure. Most profoundly, they may help reframe questions themselves. By surfacing degeneracies, highlighting hidden assumptions, or uncovering structural parallels across systems, AI can extend physics' vocabulary, creating space for new abstractions and new ways to reason.

Realizing this vision will require trust, transparency, and shared intuition between people and models. The challenge is not simply to make AI accurate but to make its reasoning legible, ensuring that it participates in physics' culture of approximation, verification, and debate. Done well, this partnership could make physics more exploratory: a discipline whose models do not just describe the world but help us ask better questions about it.

**A**rtificial intelligence is already woven into how we run experiments, simulate systems, and analyze data. The challenge ahead is to make these successes easier to sustain and build upon. The most valuable systems in the near term will amplify human judgment, helping us sift through torrents of data, accelerate simulations, and make instruments more responsive to real-time context. Their strength lies in preserving the link between data and the conditions under which those data were made.

The next stage will continue to require cultural fluency as much as technical progress. Physicists will need to learn the languages of learning; machine-learning researchers will need to respect the values that keep physics resilient to hype, including transparency, testability, and the habit of doubt. These dual responsibilities, linking technical, cultural, and infrastructural progress, mirror the themes identified in the NSF workshop's recent community paper, which calls for sustained collaboration between domain scientists and AI researchers to build trustworthy, interoperable systems.<sup>25</sup>

It also means recognizing that shared infrastructure and composable models are as essential as headline-grabbing results. Progress will depend on quieter forms of work: building infrastructure that connects modeling with measurement, designing experiments that yield training signals as well as insights, and ensuring that physical context is carried forward rather than lost. The promise of AI in science is beyond prediction; it is understanding – revealing why systems behave as they do and how their patterns connect across scales.

In physics, understanding has always come from abstraction: finding the right variables, effective models, and questions to make the complex tractable. AI may expand this vocabulary. Architectures that encode invariances, workflows that reflect physical constraints, and interfaces that expose reasoning all extend what physicists can express and test.

Better tools help us ask better questions. Structure shapes answers and guides what we consider to be askable in the first place. Some questions fail because they rest on the wrong assumptions, as when prequantum physics imagined electrons orbiting nuclei or when relativity overturned the ether. Good tools can reveal those hidden assumptions and help us reframe inquiry itself.

If we get this right, AI will broaden our reach for what we can ask, discover, and know. It will help link ideas across subfields, lower the barriers between experiment, computation, and theory, and make room for richer, testable abstractions. The goal remains the same: to make the complexity of the universe knowable to the human mind.

---

#### AUTHOR'S NOTE

The author thanks Jane Halpern and Elyssa Hofgard; Professors Jesse Thaler, Phil Harris, and Sean Lubner; and Drs. Ryley McConkey and Emily Oliphant for their contributions to the structure and framing of this essay. The author also benefited from discussions with the members of the Atomic Architects research group and at the MIT Schwarzman College of Computing ML-Driven Discovery Retreat.

This work reflects insights gained through collaborations with advisors, colleagues, and students across physics and machine learning, and through research conducted at national laboratories and their user facilities, including Fermilab, CERN, Argonne National Laboratory, the National High Magnetic Field Laboratory, and Lawrence Berkeley National Laboratory. The author is a member of the NSF Institute for Artificial Intelligence and Fundamental Interactions (IAIFI) and the NSF Institute for Artificial Intelligence and Materials (AI-MI), and acknowledges the broader NSF AI program for its support of emerging roles for AI in physics.

## ABOUT THE AUTHOR

**Tess Smidt** is Associate Professor of Electrical Engineering in the MIT Department of Electrical Engineering and Computer Science. She is also Principal Investigator of the Research Laboratory of Electronics and of the Atomic Architects research group. She has recently published in such journals as *Nature Communications*, *Advances in Neural Information Processing Systems*, and *Physical Review Research*.

## ENDNOTES

- <sup>1</sup> Inés Gil Botella and DUNE Collaboration, “The Deep Underground Neutrino Experiment (DUNE) Program,” presented at the 42nd International Conference on High Energy Physics (ICHEP2024), Prague, Czech Republic, July 18–24, 2024; and A. Weltman, Philip Bull, Stefano Camera, et al., “Fundamental Physics with the Square Kilometre Array,” *Publications of the Astronomical Society of Australia* 37 (2022), <https://doi.org/10.1017/pasa.2019.42>.
- <sup>2</sup> Benjamin P. Abbott, Richard Abbott, Thomas D. Abbott, et al., “Observation of Gravitational Waves from a Binary Black Hole Merger,” *Physical Review Letters* 116 (2016): 061102.
- <sup>3</sup> Jack Dongarra and David Keyes, “The Co-Evolution of Computational Physics and High-Performance Computing,” *Nature Review Physics* 6 (2024): 621.
- <sup>4</sup> Laura Boggia, Carlos Cocha, Fotis Giasemis, et al., “Review of Machine Learning for Real-Time Analysis at the Large Hadron Collider Experiments ALICE, ATLAS, CMS and LHCb,” arXiv (2025), <https://doi.org/10.48550/arXiv.2506.14578>.
- <sup>5</sup> Abbott, Abbott, Abbott, et al., “Observation of Gravitational Waves from a Binary Black Hole Merger.”
- <sup>6</sup> Benjamin P. Abbott, Richard Abbott, Thomas D. Abbott, et al., “Characterization of Transient Noise in Advanced LIGO Relevant to Gravitational Wave Signal GW150914,” *Classical and Quantum Gravity* 33 (13) (2016): <http://doi.org/10.1088/0264-9381/33/13/134001>; and Anamaria Effler, Robert M. Schofield, Valery V. Frolov, et al., “Environmental Influences on the LIGO Gravitational Wave Detectors during the 6th Science Run,” *Classical and Quantum Gravity* 32 (3) (2015), <http://doi.org/10.1088/0264-9381/32/3/035017>.
- <sup>7</sup> Jonas Buchli, Brendan Tracey, Tomislav Andric, et al., “Improving Cosmological Reach of a Gravitational Wave Observatory Using Deep Loop Shaping,” *Science* 389 (6764) (2025), <https://doi.org/10.1126/science.adw1291>; and Chris Whittle, Ge Yang, Matthew Evans, and Lisa Barsotti, “Machine Learning for Quantum-Enhanced Gravitational-Wave Observatories,” *Physical Review D* 108 (2023), <https://doi.org/10.1103/PhysRevD.108.043034>.

- <sup>8</sup> Ethan Marx, William Benoit, Alec Gunny, et al., “Machine-Learning Pipeline for Real-Time Detection of Gravitational Waves from Compact Binary Coalescences,” *Physical Review D* 111 (2025), <https://doi.org/10.1103/PhysRevD.111.042010>.
- <sup>9</sup> Deep Chatterjee, Ethan Marx, William Benoit, et al., “Rapid Likelihood Free Inference of Compact Binary Coalescences Using Accelerated Hardware,” arXiv (2024), <https://doi.org/10.48550/arXiv.2407.19048>.
- <sup>10</sup> Gravitational-Wave Candidate Event Database (GraceDB), “S250830m,” submitted August 30, 2025, <https://gracedb.ligo.org/superevents/S250830m/view>.
- <sup>11</sup> Mario Krenn, Yehonathan Drori, and Rana X. Adhikari, “Digital Discovery of Interferometric Gravitational Wave Detectors,” *Physical Review X* 15 (2025), <https://doi.org/10.1103/PhysRevX.15.021012>.
- <sup>12</sup> Bruce Denby, “Neural Networks and Cellular Automata in Experimental High Energy Physics,” *Computer Physics Communications* 49 (3) (1988), [https://doi.org/10.1016/0010-4655\(88\)90004-5](https://doi.org/10.1016/0010-4655(88)90004-5); Leif Lönnblad, Carsten Peterson, and Thorsteinn Rögnvaldsson, “Finding Gluon Jets With a Neural Trigger,” *Physical Review Letters* 65 (1990): 1321, <https://doi.org/10.1103/PhysRevLett.65.1321>; Bruce Denby and Stephan L. Linn, “Spatial Pattern Recognition in a High Energy Particle Detector Using a Neural Network Algorithm,” *Computer Physics Communications* 56 (3) (1990): 293–297, [https://doi.org/10.1016/0010-4655\(90\)90015-S](https://doi.org/10.1016/0010-4655(90)90015-S); and Carsten Peterson, Thorsteinn Rögnvaldsson, and Leif Lönnblad, “JETNET 3.0—A Versatile Artificial Neural Network Package,” *Computer Physics Communications* 81 (1–2): 185–220, [https://doi.org/10.1016/0010-4655\(94\)90120-1](https://doi.org/10.1016/0010-4655(94)90120-1).
- <sup>13</sup> Boggia, Cocha, Giasemis, et al., “Review of Machine Learning for Real-Time Analysis at the Large Hadron Collider Experiments ALICE, ATLAS, CMS and LHCb”; and Allison M. Deiana, Nhan Tran, Joshua Agar, et al., “Applications and Techniques for Fast Machine Learning in Science,” arXiv (2021), <https://doi.org/10.48550/arXiv.2110.13041>.
- <sup>14</sup> Boggia, Cocha, Giasemis, et al., “Review of Machine Learning for Real-Time Analysis at the Large Hadron Collider Experiments ALICE, ATLAS, CMS and LHCb.”
- <sup>15</sup> Deiana, Tran, Agar, et al., “Applications and Techniques for Fast Machine Learning in Science.”
- <sup>16</sup> TaMaryn Waters, “It’s So Hot in Tallahassee, Even the Power Hungry FSU MagLab Needed to Take a Break,” *Tallahassee Democrat*, July 31, 2025, <https://www.tallahassee.com/story/news/2025/07/31/tallahassee-mag-lab-magnetic-power-bill-fsu-extreme-heat-experiment-florida/85445060007>; and National High Magnetic Field Laboratory, “By the Numbers,” <https://nationalmaglab.org/about-the-maglab/facts-figures/by-the-numbers> (accessed October 20, 2025).
- <sup>17</sup> Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, et al., “Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation,” *APL Materials* 1 (1) (2013), <https://doi.org/10.1063/1.4812323>; Raghunathan Ramakrishnan, Pavlo O. Dral, Matthias Rupp, and O. Anatole von Lilienfeld, “Quantum Chemistry Structures and Properties of 134 Kilo Molecules,” *Scientific Data* 1 (2014): 140022, <https://www.nature.com/articles/sdata201422>; and Lowik Chanussot, Abhishek Das, Siddharth Goyal, et al., “Open Catalyst 2020 (OC20) Dataset and Community Challenges,” *ACS Catalysis* 11 (10) (2021), <https://pubs.acs.org/doi/10.1021/acscatal.0c04525>.
- <sup>18</sup> Kurt Lejaeghere, Gustav Bihlmayer, Torbjörn Björkman, et al., “Reproducibility in Density Functional Theory Calculations of Solids,” *Science* 351 (6280) (2016), <https://doi.org/10.1126/science.aad3000>.

- <sup>19</sup> Benediktus Madika, Aditi Saha, Chaeyul Kang, et al., “Artificial Intelligence for Materials Discovery, Development, and Optimization,” *ACS Nano* 19 (30) (2025): 27116–27158.
- <sup>20</sup> Jonathan Carter, John Feddema, Doug Kothe, et al., *Advanced Research Directions on AI for Science, Energy, and Security: Report on Summer 2022 Workshops* (U.S. Department of Science, Office of Scientific and Technical Information, 2023), <https://www.anl.gov/sites/www/files/2023-06/AI4SESReport-2023-v5.pdf>.
- <sup>21</sup> Ryan Jacobs, Dane Morgan, Siamak Attarian, et al., “A Practical Guide to Machine Learning Interatomic Potentials—Status and Future,” *Current Opinion in Solid State Materials Science* 35 (2025), <https://doi.org/10.1016/j.cossms.2025.101214>.
- <sup>22</sup> Alexander Bogatskiy, Sanmay Ganguly, Thomas Kipf, et al., “Symmetry Group Equivariant Architectures for Physics,” paper submitted to the Proceedings of the U.S. Community Study on the Future of Particle Physics (Snowmass 2021), March 14, 2022, <https://arxiv.org/pdf/2203.06153>.
- <sup>23</sup> Andrew Ferguson, Marisa LaFleur, Lars Ruthotto, et al., “The Future of Artificial Intelligence and the Mathematical and Physical Sciences (AI+MPS),” arXiv (2025), <http://arxiv.org/abs/2509.02661>.
- <sup>24</sup> Yifan Cheng, Nikolaus Grigorieff, Pawel A. Penczek, and Thomas Walz, “A Primer to Single-Particle Cryo-Electron Microscopy,” *Cell* 161 (3) (2015); and Allison Doerr, “Cryo-Electron Tomography,” *Nature Methods* 14 (2017), <https://doi.org/10.1038/nmeth.4115>.
- <sup>25</sup> Ferguson, LaFleur, Ruthotto, et al., “The Future of Artificial Intelligence and the Mathematical and Physical Sciences (AI+MPS).”