The scientist in the machine: how AI is rewiring scientific research, from protein folding to autonomous labs

Francesco Barbato · Full Stack · Cybersecurity · AI

Published 26 May 2026 · 11 min read

ABSTRACT

In 2024 two Nobel Prizes went to work rooted in machine learning. A technical look at how AI is changing the scientific method: where it works (AlphaFold, GNoME, self-driving labs), where it fails (hallucinated citations in the literature) and what it still can't do.

Keywords: AI · Machine learning · Scientific research · Deep learning · AlphaFold · Drug discovery · Materials science · Research integrity

In October 2024 the Royal Swedish Academy of Sciences did something it had not done before: in the same year it awarded Nobel Prizes in two different disciplines for work rooted in machine learning. The Physics prize went to John Hopfield and Geoffrey Hinton for the foundations of artificial neural networks. The Chemistry prize went to Demis Hassabis and John Jumper for AlphaFold, and to David Baker for computational protein design.

It was not a symbolic gesture. It was an acknowledgement that the methodology of natural science (hypothesis, experiment, observation, refinement) is being structurally rewritten by systems that can search hypothesis spaces orders of magnitude faster than humans. This article looks at how that is actually happening, where it is working, where it is failing and what the next few years probably hold.

1 The protein problem, and why it broke open

For roughly fifty years, predicting the three-dimensional structure of a protein from its amino acid sequence was one of biology's most expensive open problems. The experimental approaches, X-ray crystallography and cryo-electron microscopy, could take months or years per protein and needed physical samples that didn't always crystallise.

In 2020 DeepMind's AlphaFold 2 changed that. For most targets it predicted structures with accuracy comparable to experiment, and in 2021 the code was released together with a searchable database that today holds predicted structures for some 200 million proteins. According to the Nobel

Committee, within four years more than two million people from 190 countries had used it, and the main paper has collected tens of thousands of citations.

> **WHY THIS MATTERS TECHNICALLY**
>
> AlphaFold 2 combined transformer-style attention over multiple sequence alignments (MSAs) with a geometric "structure module" that predicted the 3D coordinates of the protein end to end. Because the whole pipeline was differentiable, the network could learn protein geometry the way image classifiers learn what a cat looks like.

In May 2024 DeepMind and Isomorphic Labs presented AlphaFold 3, which predicts the structure of proteins, DNA, RNA, ligands and their interactions, with the explicit goal of speeding up drug discovery; code and weights for academic use followed in November 2024. The jump from 2 to 3 is not trivial: predicting how a drug-like small molecule docks into a protein's binding pocket is a different and harder problem than predicting the shape of the protein alone, because it means modelling the joint conformational landscape of two interacting species.

2 From folding to designing: the inverse problem

Predicting structure from sequence is one direction. The other, designing a sequence that folds into a desired structure with a desired function, is what David Baker's group at the University of Washington has worked on for decades, developing many of the algorithms and tools the community uses.

The practical consequence is that researchers can now specify something like "a protein that binds this viral spike with very high affinity and is stable above 60 °C", and AI systems will propose candidate sequences for experimental validation. Custom enzymes, vaccine immunogens and binders for previously "undruggable" targets increasingly come out of RFdiffusion-class models (Watson et al., *Nature* 620, 1089–1100, 2023) and their successors.

3 Materials science: centuries of discovery in one paper

In November 2023 DeepMind published a result that did for inorganic materials what AlphaFold had done for proteins. GNoME (Graph Networks for Materials Exploration) predicted 2.2 million new crystal structures, which DeepMind compared to roughly 800 years' worth of knowledge, and about 380,000 of them were assessed as stable enough to be candidates for synthesis.

The scaffolding is worth understanding. GNoME uses graph neural networks in which nodes are atoms and edges are bonds, a natural representation for crystalline solids. The team used active learning: the model proposes candidate crystals, those are checked with Density Functional Theory (DFT) calculations and the resulting high-quality data goes back into training. Over the rounds, the precision of its stability predictions rose from around 50% to about 80%.

For context, the number of known stable materials grew almost tenfold, to roughly 421,000. Among the results were 528 promising lithium-ion conductors, relevant to better batteries.

> **THE SYNTHESIS GAP**

> A predicted stable material is not a synthesised one, and even after new materials are made it usually takes decades for industry to bring them to market. The bottleneck has moved from "what should we try?" to "can we actually make it?", which is exactly where the next wave is happening. Caution is also due on the claims themselves: independent materials scientists have questioned how many of the "new" compounds announced by this line of work, including the autonomous-lab synthesis results, are genuinely novel and useful.

4 Self-driving laboratories: closing the loop

If AI can propose hypotheses faster than humans can test them, the obvious next step is to automate the testing too. That is the premise of self-driving laboratories (SDLs), which in recent years have moved from prototypes toward real use.

SDLs are commonly classified by five levels of autonomy: Level 1 (assisted operation, such as robotic liquid handlers), Level 2 (partial autonomy, such as protocol generation), Level 3 (conditional autonomy: at least one full cycle of the scientific method done autonomously), Level 4 (high autonomy: protocol generation, execution, analysis and hypothesis adjustment) and Level 5 (full autonomy, an AI researcher, not yet achieved).

Table 1. Levels of autonomy of self-driving laboratories, with example systems from the literature. Where a given system sits is partly a matter of definition.

Autonomy level	Capability	Example systems
Level 1	Machine-assisted tasks	Liquid handlers, data analysis software
Level 2	Protocol generation	Aquarium
Level 3	One full scientific-method cycle	iBioFab, Mobile Robot Chemist
Level 4	Hypothesis-driven autonomous research	A-Lab, Adam, Eve, MicroCycle
Level 5	Full AI researcher	Not yet achieved

The state of the art sits at the upper levels, in well-defined chemistries. Lawrence Berkeley National Laboratory's A-Lab combines literature data, machine learning and active learning to plan and synthesise target compounds, run X-ray diffraction and interpret the results. At North Carolina State University, a self-driving lab demonstrated a technique that collects data at least ten times more efficiently than previous approaches, speeding up materials discovery while cutting cost and environmental impact.

Large language models have started to act as the "cognitive layer" of these systems. Boiko et al. proposed Coscientist, an LLM-driven system able to design, plan and control robotic operations for chemical experiments: it reads the literature, proposes synthesis routes, instructs a robotic platform, interprets spectroscopic results and updates its plan.

5 Medical research and clinical AI

Beyond the molecular scale, AI has become a front-line tool in clinical research and diagnostics. In 2025 Microsoft's Diagnostic Orchestrator (MAI-DxO) solved complex medical cases with 85.5% accuracy, against about 20% for the physicians it was compared with. The number needs careful unpacking: the cases came from *New England Journal of Medicine* clinicopathological conferences, which are deliberately hard, and the physicians worked without access to colleagues or reference materials. Even with those caveats, a difference of that size on the hardest cases is significant.

The framing is shifting too. Peter Lee, president of Microsoft Research, predicts that in 2026 AI will not just summarise papers, answer questions and write reports but actively join the process of discovery in physics, chemistry and biology: generating hypotheses, using tools that control experiments and collaborating with human and AI colleagues. It is a prediction, not a result.

6 The dark side: hallucinations enter the permanent record

The same systems that propose protein structures and synthesis routes also fabricate citations with uncanny confidence. By 2026 the problem is no longer anecdotal.

In January 2026, GPTZero analysed the 4,841 papers accepted at NeurIPS 2025, the premier AI conference, and found at least 100 confirmed hallucinated citations across 51 papers, about 1% of the accepted work, despite review by several experts per paper.

It is not limited to AI conferences. An audit by Columbia University's nursing school of 2.5 million open-access biomedical papers found 4,046 fabricated citations in 2,810 papers published between January 2023 and February 2026. And a late-2025 survey by the publisher Frontiers found that more than half of reviewers are already using AI tools, often outside their publishers' guidelines.

The correction machinery is not keeping up: in that audit, 98.4% of the affected papers had received no action from their publisher. Once a fabricated citation is in the literature it tends to stay there, and later papers can cite the one that contains it, propagating the error.

> **THE REPRODUCIBILITY COMPOUND**
>
> This is not a parallel crisis to the existing reproducibility crisis; it accelerates it. A rising flood of unreliable AI-generated content now permeates the academic literature, and fabricated citations break the evidence chain: if a result rests on a study that doesn't exist, it can't be replicated even in principle.

7 The methodology of the next decade

Several technical trends are converging into what looks like a new scientific methodology.

Hypothesis generation as search. The classical model, in which a human forms a hypothesis from intuition and reading, is being supplemented (not replaced) by systems that explicitly enumerate and score hypotheses over vast combinatorial spaces. GNoME's 2.2 million crystals are one form of

this; AlphaFold's 200 million predicted structures are another. The human role shifts from generator to curator.

Active learning loops. The training data for these models is increasingly generated by the models themselves, validated by simulation or experiment and fed back into training. It is a closed loop that improves with use, and fundamentally different from the static-dataset paradigm of the first decade of deep learning.

Foundation models for science. General pretrained models, analogous to GPT-class language models but trained on scientific data (sequences, structures, spectra, equations), are starting to displace task-specific architectures. Much of the progress now happens in post-training, where models are refined on specialised data, which favours open-source models that can be customised and fine-tuned for specific applications.

Symbolic-statistical hybrids. Pure neural networks struggle with rigorous derivation and provable correctness. Researchers at the University of Pennsylvania introduced “Mollifier Layers”, a technique that integrates classical mathematical smoothing functions into neural networks to solve inverse partial differential equations with far greater stability, a long-standing problem when high-order derivatives are computed on noisy real-world data. Applications span genomics, materials science, climate modelling and chromatin biology.

8 What AI is not (yet) doing

The hype gradient on this topic is steep, so it is worth listing what is still out of reach.

AI does not formulate genuinely new theoretical frameworks. It has not derived a new physical law, proposed a new symmetry principle or unified two separate fields. It accelerates work within existing frameworks. The leap from Galilean mechanics to general relativity is not, currently, the kind of leap a transformer makes.

AI does not, on its own, run frontier wet-lab experiments. Self-driving labs work in well-characterised chemistries and well-defined parameter spaces. The truly exploratory experiment, where you don't know what equipment to use, what variables to measure or what failure looks like, remains a human activity.

AI does not understand causation. It detects correlations exquisitely well. Telling “X causes Y” from “X correlates with Y because both are caused by Z” still takes interventional reasoning, controlled experiments and human judgement about which counterfactuals are worth running.

> **A SOBER NOTE**
>
> Current and planned spending on data centres is the largest technology project in history, and some observers speak of a bubble: revenues are underwhelming, the gains from ever larger language models seem to be levelling off, and there are theoretical limits on how efficiently they can learn straightforward concepts, as Stuart Russell of UC Berkeley has argued. Progress in scientific AI is real, but it is not unbounded.

9 Conclusion: the scientist's new role

The headline finding of the last five years is not that AI replaces scientists. It is that AI changes which parts of science are scarce. Hypothesis enumeration, literature synthesis and pattern recognition in large datasets are no longer rate-limiting. What stays scarce is judgement: deciding which of a million predicted crystals to actually synthesise, which fabricated citation to catch before it propagates, which clinical AI recommendation to override, which new experimental approach to invent because no model has thought of it yet.

The Nobel Committee's 2024 decision recognised not just specific tools but a methodological shift, with consequences for medical research, biotechnology and materials science. The next ten years of science will be measured, in significant part, by how well the field integrates these tools while preserving the epistemic standards that made science worth doing in the first place.

Primary references: Jumper et al., Nature 596, 583–589 (2021); Tunyasuvunakool et al., Nature 596, 590–596 (2021); Abramson et al., Nature 630, 493–500 (2024); Merchant et al. (GNoME), Nature 624 (2023); Watson et al., Nature 620, 1089–1100 (2023); the 2024 Nobel Prize press releases (NobelPrize.org); the Stanford HAI AI Index; and reporting from Nature, MIT Technology Review, Fortune and TechCrunch through May 2026. Figures on hallucinated citations come from GPTZero's NeurIPS 2025 analysis and from the Columbia University School of Nursing audit.

HOW TO CITE THIS ARTICLE

Francesco Barbato (2026). The scientist in the machine: how AI is rewiring scientific research, from protein folding to autonomous labs. localhost:3000.

Online version: http://localhost:3000/en/blog/ai-scientific-research-deep-dive