Introductory Notes: On AI, Science & the Future of Discovery

James Manyika

The last few years have seen a flurry of major scientific awards at the intersection of AI and science: In 2024, the Nobel Prize in Physics was awarded to Geoffrey Hinton and John Hopfield for their pioneering work on machine learning with artificial neural networks, which has formed the basis of modern AI, while the Nobel Prize in Chemistry recognized David Baker, Demis Hassabis, and John Jumper for their work on protein design and structure prediction. In 2025, the Nobel Prize in Physics went to John Clarke, Michel Devoret, and John Martinis for laying the foundations of quantum computing and quantum AI. The Turing Award ("the Nobel for Computer Science") was conferred in 2024 on Andrew Barto and Richard Sutton for their contributions to reinforcement learning and, not long before that, in 2018 on Yoshua Bengio, Geoffrey Hinton, and Yann LeCun for their pioneering work on deep learning. In addition to the achievements, these awards – spanning multiple scientific disciplines and drawing on ideas from the different disciplinary backgrounds of the recipients – underscore the bidirectional nature of progress at the intersection of AI and science: science advancing AI and AI advancing science.

The work by Baker and by Hassabis and Jumper is notable not only for *what* was achieved but also *how* it was achieved – and for what *that* portends for the future of discovery.[1]

With AlphaFold, Hassabis and Jumper solved the fifty-year grand challenge of protein-structure prediction. In his 1972 Nobel lecture, biochemist Christian Anfinsen had postulated that a protein's structure is fully determined by its amino acid sequence. But predicting this shape computationally was nearly impossible because a typical protein has an astronomical number of possible configurations, far beyond what could simply be searched by brute force. By achieving experimental accuracy, AlphaFold demonstrated that AI could learn the "grammar" of biology from data, bypassing the need for computationally expensive *ab initio* simulations. The team released the predicted structures of over two hundred million proteins – nearly the entire known protein universe – which have now been used

https://doi.org/10.1162/DAED.e.970

by more than 3.3 million scientists in over 190 countries. Baker's work asked the inverse question : Can we design a protein that binds to a specific target ? He and his team adapted AI diffusion models to start with random amino acid coordinates and then "denoise" them to form a valid protein backbone that fits a specific constraint. For example, if a researcher wants a protein that binds to a specific site on a virus, they condition the diffusion model on that site and the model "hallucinates" (a useful phenomenon in this case) a protein shape that perfectly complements it.[2] In experimental validation, the success rate of proteins that bind skyrocketed by orders of magnitude compared with previous methods, making possible the design of proteins that were formerly impossible to engineer.

To these AI-enabled advances we could add many more from just the last few years in genomics, neuroscience, health care, chemistry, materials science, plasma physics, astrophysics, climate science, and beyond.[3] To be clear, the now constant stream of announced advances has not been without instances of hyperbole, overstatement, false claims, *in silico* (computer simulated) breakthroughs that don't match real-world reality, contested results, and outright failures. In this regard, AI is like every paradigm change that preceded it.

The combination of progress in AI and its related techniques, its expanding and general usefulness in science, and its celebrated landmark contributions to discovery has led in recent years to recognition by the scientific community that a new era of AI for science has arrived. Indeed, many universities, research consortia, national academies, national laboratories, and governments have launched AI-for-science initiatives.[4] In the 2023 *Nature* article "Scientific Discovery in the Age of Artificial Intelligence," a group of prominent scientists put the matter succinctly : AI has now transcended its role as a peripheral tool for analysis to become a foundational mechanism for discovery.[5]

Yet there are also challenges and an unease about the shifts accompanying this new era : for example, that academic and national laboratories are ill-suited for the "data-intensive and compute-intensive training regime of AI models" and face a brain drain to the private sector, where most frontier AI research and AI-enabled science now takes place (though often involving academic collaborators). Many have raised concerns about AI's impact on *how* science is done, from reproducibility and contamination of the scientific literature to epistemic challenges and expanded safety and ethical considerations.[6] Nevertheless, most scientists acknowledge that AI-enabled science is here to stay and that AI should be harnessed to advance science, perhaps with a broader or reimagined view of *what* could be achieved and *how*.

So, what is the future of scientific discovery in the age of AI ?

The most recent volume of *Dædalus* that focused on artificial intelligence was published in spring 2022. That volume, which I had the honor of editing, focused

on AI & Society. It was not the first *Dædalus* issue to focus on AI – that distinction goes to a volume published in 1988. The 1988 *Dædalus* volume captures the tension at the time: on one hand, disappointment that AI had not lived up to the high expectations set at the field's founding in the 1950s (typified by Hilary Putman's essay "Much Ado About Not Very Much"); on the other, a sense that the field was on the verge of a rebirth. The 2022 volume explores the renewed technical progress in AI as well as its implications for society, from jobs and the economy to democracy and distrust, governance, national security, justice, and ethics.

Much has happened since the publication of that 2022 volume of *Dædalus*: It came out six months before the launch of ChatGPT demonstrated to the public just how far AI had come. Progress in AI has continued at a rapid pace since, as chronicled in numerous reports, such as our annual Stanford AI Index reports in this period (2022–2026).[7] With this progress, its intersections with society have expanded, resulting in more excitement about AI's useful and beneficial possibilities as well as increased concern about the challenges and risks. The stakes – positive and negative – have grown in both directions: from AI's impact on what individuals are now able to do, to its effects on productivity and growth in the economy, and on jobs, education, and creativity; to its use, misuse, missed use, and the societal implications; to its infrastructure and energy footprint, its geopolitics, its regulation.[8] These developments warrant deeper assessment and discussion. But this is not that volume; this volume is focused on AI and science.

For science, this moment is not unlike that before the arrival of ChatGPT in 2022. Those at the dual frontiers of AI and its use in science recognize how AI is advancing science, but the general public remains much less aware. This gap stems partly from the indirect nature of the impact on individuals and society of AI's contributions to science – unlike the firsthand experiences that many people have with chatbots like Claude, ChatGPT, or Gemini – and also from the topicality and, some would argue, immediacy of AI's other aforementioned impacts. When it comes up at all, the prospect of AI advancing science is often touted as something to look forward to, a reason to take AI seriously, the pot of gold at the end of the rainbow. That view is correct in that many more of the possibilities of AI for science lie in the future, but it misses its ongoing contributions to science and the advances *already* being achieved, as well as the extent to which society is already benefiting in areas ranging from health care to weather forecasting to food security.

While this *Dædalus* volume may help address gaps in awareness, that is not its primary aim. Rather, its main motivation is to explore the future of scientific discovery in the age of AI. Given the unwieldy scope this suggests, this collection focuses on and draws primarily from scientists working at the frontier of the intersection of AI and science, while acknowledging that this does not represent the full range of activities or views in the scientific community (nor does it cover all the issues that our core question conjures).

The volume consists of contributions from thirty-three scientists, spanning the life sciences and medicine, cognitive science and neuroscience, the physical and earth sciences, chemistry and materials science, computer science, mathematics, and the social sciences. Their contributions are organized into five sections : The first engages two notable scientists at the frontier of AI's advance. The second section, "AI in the Life Sciences & Physical Sciences," explores what's being achieved and what lies ahead. The third, "The Science of AI & Developing AI for Science," examines AI's current limitations and efforts to advance it to further enable discovery. The fourth, "The Social Science of AI for Science," explores the wider implications of AI-assisted science, including on how science is done, the role of the scientist, the scientific method, and understanding. The final section consists of "personal briefs" from early-career scientists for whom the question "Are you an AI researcher or a scientist ?" feels outdated, as well as a statement by an artist on "When AI Meets Art at the Scales of Science." Despite these groupings, each contribution should be read as a unique response to this volume's motivating question. I am grateful to the contributors for accepting my invitation to participate in this volume.

As when I edited the 2022 *Dædalus* issue on AI & Society, I am not a bystander – not in AI generally nor its intersection with science specifically. I lead research teams in these areas, have coauthored related papers (some cited in this volume), and am involved in relevant work in academia. Indeed, my research in AI started in the 1990s and focused on solving scientific questions. Undoubtedly, this volume is colored not only by my various involvements but, most importantly, by the views of its contributors. To supplement these perspectives, I have included extensive endnotes to facilitate further exploration.

Before proceeding, it may be useful to define AI. As noted in my introduction to the 2022 *Dædalus*, the field has a historical grounding in the now-famous 1956 Dartmouth Summer Research Project on Artificial Intelligence, to which the term "artificial intelligence" is largely attributed.[9] In that volume, I also noted what Marvin Minsky had earlier pointed out : that AI had become a "suitcase word" packed variously with a wide range of concepts and techniques.[10] That problem has since worsened. In our context of AI and science, the term encompasses a Minsky-ean suitcase of approaches, from statistical, computational, and machine-learning and deep-learning techniques to analyze, infer, and discern patterns, learn from data, form abstractions of sensory inputs, build models of natural phenomena, and make predictions, to approaches and systems that simulate physical laws, explore possibility spaces, generate hypotheses, propose designs, and steer experiments, including through the human-like capabilities (such as reasoning) that those at Dartmouth had in mind.

It is sometimes assumed that the intersection of AI and science is something that took-off with the arrival of large language models (LLMs). Therefore, a brief history that selectively signposts the lineage of a few ideas that underpin and continue to animate this intersection may be useful.

For centuries, scientific discovery rested on two pillars: *theoretical science*, rooted in the deductive logic of mathematics and natural philosophy, seeking to construct models of reality – from Newton's laws to Einstein's equations – with predictive power; and *empirical science*, grounded in observation and experimentation serving as the arbiter of a theory's validity. The latter half of the twentieth century saw the crystallization of a third pillar: *computational science*.[11] This did not merely automate arithmetic and extend the power of the former two; it brought forward a fundamentally distinct methodology for investigating phenomena that were analytically intractable, experimentally inaccessible, or both.[12] While this may have crystallized midcentury with the advent of digital computers, the history of computational science stretches back to antiquity (with much of the computation painstakingly human, occasionally mechanical).[13] That history is bidirectional: computational approaches advancing science and science advancing computational approaches, with pioneering scientists from multiple disciplines operating on this two-way street right up to the present with AI.

The polymaths Alan Turing and John von Neumann stand out in this lineage of ideas.[14] Among its various insights, Turing's seminal 1936 paper "On Computable Numbers" suggests that if reality is based on scientific laws and those laws are computable, then reality could be modeled as a Turing machine and the nature of reality could be informational – an idea espoused by Demis Hassabis and some others.[15] Turing later proposed that a network of randomly connected nodes (like those in the brain) could be trained rather than programmed and could organize itself for a definite purpose. He then inverted the problem: Instead of using biology to inspire AI, how could the computer be used to elucidate a biological question? He showed that biological patterns could be explained by the interaction of simple chemical substances, which he modeled using a reaction-diffusion model. Taking a different tack, in his 1958 paper "The Computer and the Brain," von Neumann compares the artificial and the natural automaton, concluding that the computer (artificial) was high-speed and high-precision, while the brain (natural) was a slow, low-precision mixed system that was probabilistic and massively parallel. Motivated by the question of how complex organisms evolve from simpler ones without violating the second law of thermodynamics, he introduced in 1966 his "Theory of Self-Reproducing Automata." In the 1980s, computer scientist and physicist Stephen Wolfram would develop this further and show that simple automata offered a computational alternative to the differential equations that had dominated physics.[16]

In the wake of Turing and von Neumann, the pursuit of AI branched in the 1950s into two competing visions – the "symbolic" (mind) approach and the "connectionist" (brain) approach – with science a primary proving ground. The symbolic approach, championed by Allen Newell, Herbert Simon, and Marvin Minsky, defined intelligence as symbol manipulation according to formal rules and heuristics. In 1956, their Logic Theorist algorithm demonstrated that a machine could

prove mathematical theorems – specifically, thirty-eight theorems from *Principia Mathematica* – using heuristic search rather than brute force.[17] This suggested that if scientific knowledge could be axiomatized, machines could theoretically derive all consequences of those axioms and essentially compute the truths of the universe. On the other branch, in 1951, Marvin Minsky and Dean Edmunds built SNARC, the first connectionist neural network to demonstrate reinforcement learning; it simulated a rat navigating a maze. And in 1957, Frank Rosenblatt developed the perceptron, a probabilistic model capable of learning from data. However, such learning models were overshadowed by the success of general problem solvers built on heuristics, logic, and knowledge of science.

By the 1970s, as it became apparent that general problem solvers could not scale to complex scientific domains in which the combinatorial explosion overwhelmed simple logic, AI had already begun to shift to domain-specific approaches. For example, Margaret Dayhoff's key insight that biology was inherently an information science allowed her to pioneer probabilistic models for comparing protein sequences, enabling the construction of phylogenetic trees and the study of molecular evolution. DENDRAL, started in 1965 by Edward Feigenbaum, Joshua Lederberg, and Bruce Buchanan, used heuristics derived from chemists to analyze mass spectrometry data and elucidate molecular structures. Its successor, Meta-DENDRAL, became the first to automate discovery by inferring new cleavage rules for mass spectrometry from data. But by the late 1980s, such expert systems were hitting a wall. Extracting rules from experts was slow and expensive, and the systems couldn't handle edge cases. Simultaneously, scientific datasets were growing too large for manual rule creation but were perfect for probabilistic analysis.[18]

Meanwhile, the field was rediscovering the learning approach to AI. In 1982, John Hopfield developed a model of associative memory by drawing an analogy to magnetic spin systems in condensed matter physics and statistical mechanics. The Hopfield model explained how networks of neurons could store and recall facts and led to the creation of Boltzmann machines, thereby connecting statistical mechanics with machine learning. This influenced Geoffrey Hinton's study of Boltzmann machines; thus, the progression from Hopfield models to Boltzmann machines and belief networks became the "arc of discovery" to deep learning, as Alán Aspuru-Guzik and Surya Ganguli discuss in this volume. Around the same time, while exploring how neural processing works in the brain, David Rumelhart, Geoffrey Hinton, and Ronald Williams were using back-propagation to train neural networks. They cited Yann LeCun, who by 1989 was applying back-propagation to handwritten zip code recognition, a precursor to machine vision. It was also during this period that Andrew Barto and Richard Sutton laid the foundations of modern reinforcement learning, drawing from fields as diverse as psychology to control theory.[19] This period represented the rebirth signaled by the 1988 *Dædalus* volume, especially in the essay by Hubert Dreyfus and Stuart Dreyfus,

"Making a Mind Versus Modeling the Brain : Artificial Intelligence Back at a Branchpoint."[20]

In the 2022 *Dædalus* volume on AI & Society, Jeffrey Dean describes the period since 2012 as a "golden decade" for AI, driven by advances in deep learning, computing (chips, infrastructure, and tools) suited to deep learning, and the explosion of data to train models.[21] In that same volume, Kevin Scott highlights work decades earlier – by Ray Solomonoff and Claude Shannon linking entropy, information, and probability theory with machine learning – that now enables deep-learning models to compress and generalize what they learn.[22] In this volume, LeCun adds additional insights on AI's progress during this period, while Aspuru-Guzik as well as Charlotte Bunne and Aviv Regev note AI's contributions to biology and chemistry.[23] AI for science in this era came to mean the use of deep neural networks as universal function approximators to learn from data and model complex scientific phenomena, make predictions, and accelerate simulation. The Nobel-worthy breakthroughs that I highlighted at the beginning became defining scientific achievements of this era.

So where are we now, and what is the future of discovery in the age of AI?

What is the perspective from the AI frontier?

Outside the field, it is sometimes assumed that those at the frontier of developing AI agree on everything, or that there is consensus on its technical and scientific trajectories. The dialogues with Nobel laureate Demis Hassabis and Turing awardee Yann LeCun that open this volume illustrate that this is not the case. In these conversations we find agreements, yes, but also differences and debates on the limits of LLMs, scaling, and tokenization, further breakthroughs needed in AI, artificial general intelligence (AGI), how AI can advance science and vice versa, and the future of discovery and open science.

While readers can engage directly with the perspectives of Hassabis, LeCun, and other contributors, it may also be useful to sample the views of other pioneering scientists to provide context for this volume and as inducement to explore further.

Dario Amodei, a biophysicist and cofounder of Anthropic, has stated that the "scaling hypothesis" is an empirical fact, that current transformer architectures, scaled sufficiently, will eventually master causal mechanisms of, for example, biology simply by observing enough biological data. Amodei is concerned, however, with interpretability and sees it as critical for scientific understanding.[24] As counterpoint, Ilya Sutskever, a pioneering researcher and cofounder of OpenAI and Safe SuperIntelligence, once an advocate of scaling, recently argued that the "age of scaling" is hitting diminishing returns and a new paradigm is needed, observing that "we are back in the age of wonder and discovery." Like some others in this volume, he advocates for System 2 approaches as critical for science, which requires deep, multistep reasoning rather than just next-token prediction.[25] Fei-Fei Li, another

pioneering computer scientist in AI, believes what's required are world models to reason about three-dimensional geometry and physics, as well as a body to interact with the physical world.[26]

Turing awardee Yoshua Bengio stands in contrast to the growing agentic vision for AI. He argues that until we solve alignment – ensuring that AI systems' goals, behaviors, and actions match human preferences and goals – giving AI agency (the ability to pursue goals in the world) is risky, even in a scientific context. He is developing "Scientist AI," a system designed to understand the world, generate explanatory theories, and answer questions, thus restricting AI to epistemic rather than instrumental tasks.[27] Stuart Russell, computer scientist and coauthor (with Peter Norvig) of the most read textbook on AI, also argues for non-agentic or "oracle" architectures, citing alignment and safety risks. Russell believes that AI can take science beyond "Edisonian" trial-and-error and move fields like biology from a descriptive science to a predictive one. Describing LLMs as "circuitry" without "concepts," he sees gaps that are critical for science.[28]

In perhaps the most provocative and historically grounded thesis, Richard Sutton, another Turing awardee, has offered what he calls "the bitter lesson": that every time researchers tried to build human knowledge into AI, it worked in the short term but blocked long-term progress; and that by encoding our current scientific theories into AI, we bias it toward what we already know and thus prevent it from discovering radically new physics or biology. Sutton argues we must instead build general learning methods, including AI systems capable of search and "meta-methods" to discover science on their own. He states that LLMs on their own are a dead end because they are static, that true scientific AI must learn and interact with the world to discover new laws.[29] It's worth noting that one of the possibilities of the learning approach is self-improving AI. On one hand, self-improvement could lead to AI that is even more capable of advancing discovery; on the other hand, if self-improvement is recursive and outruns our capacity to ensure safety, alignment, and governance, it could create unprecedented risks. While the idea of recursive self-improvement is not new, the recent progress in AI has attracted some to pursue its development, and has sparked debates about its feasibility and concerns about its safety.[30]

Such varied views suggest that while AI advances have led to the development of much that is useful, including for science, the question *What is the future of discovery?* applies equally to artificial intelligence as a scientific field itself.

How is AI advancing discovery, and what are the possibilities for the future?

In his contribution "From Alchemy to AIchemy: On Matter, Minds & Tools," Alán Aspuru-Guzik explores the transition from empirical alchemy to twentieth-century integration of computational science with chemistry and materials science to the

AI-enabled advances of today and the future.[31] He also describes the evolution of modalities of discovery with AI from tools to agents, such as *El Agente* for quantum chemistry, to self-driving laboratories with paradigm-shifting implications for scientific understanding. Using examples of advances in biology, materials science, plasma physics, and mathematics, in "Unlocking Scientific Intuition & Reasoning at Digital Speed," Pushmeet Kohli frames AI not merely as a computational accelerator but as a novel epistemic instrument that will take discovery further to predictive design, generating hypotheses and orchestrating experiments. This would precipitate a profound shift in the role of the scientist: from solver of puzzles to "architect of questions." Coupling his optimism with caution, Kohli invokes E. M. Forster's *The Machine Stops* to warn against a passive reliance on systems we cannot interpret, arguing that while AI may collapse the distance between question and answer, humanity must maintain the *why* behind the question.

With an eye on breakthroughs in diseases, Bunne and Regev, in "Beyond Representation: AI in Cellular Discovery," and Anna Greka, in "Building the Drug Discovery Engine of the Future with AI-Empowered Nodal Biology," focus on cellular biology. They articulate distinct visions in which the "cell-prediction problem" becomes the fulcrum of discovery. Bunne and Regev concentrate on the emergence of "virtual cells": computational surrogates built on AI models that integrate multi-omic and spatial modalities to simulate and explore cellular dynamics. Greka argues that while the field celebrates AI's triumphs in drug design and clinical optimization, the rate-limiting step remains target identification. She is pursuing the cell-prediction problem by forecasting responses of any human cell to perturbations using AI models trained on massive cell datasets. Such "nodal biology" could advance the identification of mechanisms across many genetic pathologies. In "From Pixels to Minds: Mapping & Understanding the Brain with AI," Viren Jain and Jeff Lichtman turn AI's microscope to the brain's architecture and the intractable complexity of mapping synaptic connectivity, where manual analysis of petascale electron microscopy data is impossible. By evolving from static anatomical catalogs to AI models that simulate neural dynamics and predict functional perturbations, they are charting a path to decoding the brain's neural syntax of healthy circuits to develop fingerprints of pathologies such as schizophrenia and autism. In an inversion reminiscent of Turing, they are also using the AI-enabled nanoscale connectomics to investigate whether the brain's neural circuits implement computational primitives analogous to those in AI systems, such as the attention mechanisms in transformers.

In "The Algorithmic Planet," Anna Michalak and John Platt reframe climate action and understanding of the physical Earth as problems of constrained optimization. They distinguish between "tractable" problems and the "intractable frontier" of long-term projection in which nonstationarity and chaotic variability confound purely data-driven approaches. The tractable frontier is where AI is increasingly being deployed around the world, for example, optimizing power grids, avoiding

contrails, and improving predictions of extreme weather events like floods, hurricanes, and wildfires.[32] With progress in Earth foundation models, Michalak and Platt weigh this revolution against the paradox of AI's growing carbon footprint – not unlike Gege Wen's "double Ouroboros" and Aspuru-Guzik's warning of a possible "AIocene."[33] Wen's double Ouroboros also suggests that to scale AI, we must reinvent energy, and to reinvent energy, we must use AI, as in her work navigating the intricate physics of the subsurface. Meanwhile in "AI Reaches for the Stars," Stella Offner details how AI has graduated from classifying celestial morphologies to accelerating computationally prohibitive simulations of galaxy evolution and beyond. Shirley Ho, in "Building an AI Polymath," and Mario Krenn and Heather Champion, in "Philosophy of Autonomous Science : Ten Questions for the Coming Age of Artificial Scientists," survey other notable advances in astrophysics. However, as Offner notes, there are profound epistemic challenges : for example, unlike in the terrestrial sciences, astronomical properties – mass, temperature, density – cannot be measured directly but must be inferred from observed light. This exacerbates the tension between scientific rigor and the "black box" opacity and predictive power of deep learning that many in this volume note. At the same time, AI may offer new avenues to tackle what some have characterized, for example, as the "crisis in particle physics," in which experimental results have failed to yield new insights for questions in the standard model. So rather than analyze data from new instruments to test and confirm existing theories, why not use AI to explore unencumbered by current theories ?[34]

What's clear is that many AI-enabled advances so far concern high-dimensional problems for which data are abundant, as opposed to areas governed by chaotic, nonstationary dynamics or those with sparse data, which for now remain the "intractable frontier."[35] Further, many examples illustrate what may be the most distinctive aspect of the AI-for-science era : the possibility of discovery at unprecedented scale, scope, and speed, such as taking on *all* proteins and cells, exploring vast combinatorial spaces (computationally out of reach with brute force methods), learning from multidimensional and multimodal data – even multidisciplinary theories and ideas – simulating dynamics and interactions, and probing and learning from perturbations *in silico*. Yet sometimes there are significant discrepancies between simulation or generated artifacts and the real world ; as Connor Coley puts it, "every [*in silico*] proposal eventually collides with the physical world." Such discrepancies, together with the inconsistency or "jaggedness" of current AI, reveal the paradox that along with novel insights there is the possibility of artifacts not grounded in reality or hallucinations that could mislead or indeed inspire novel directions (such as in Baker's Nobel-worthy work). But as Turing observed, "If a machine is expected to be infallible, it cannot also be intelligent," suggesting that for an AI (or human mind) to be capable of novel discovery, it must possess the capacity to make inductive leaps, guesses, and mistakes, and learn from them. Thus,

in the hands of a discriminating scientist or one able to validate *in silico* results in a "wet lab," this jagged AI can be usefully employed to advance science.[36]

The formal sciences – distinguished by their abstract structures and axiomatic and logic formalism rather than empirical observation of the physical world – have also been a proving ground for AI. In computer science, there have been advances on some long-standing problems: for example, in matrix multiplication, sorting and hashing algorithms, combinatorics and computational complexity, and theorem proving.[37] What about mathematics? In their recent paper "Mathematical Exploration and Discovery at Scale," Fields medalist Terence Tao and his colleagues find that while current systems like AlphaEvolve (which achieved gold medal–level performance at the International Mathematical Olympiad, as Kohli discusses) excel at "constructive mathematics at scale," they encounter epistemological boundaries and falter when problems resist formulation as smooth, optimizable functions amenable to hill-climbing.[38] This is not unlike Melanie Weber's view in her essay that AI is particularly effective for "needle in a haystack" problems in mathematics. Thus, while AI can discover complex "constructions," it currently lacks the capacity for deep new insights, serving for now as a powerful prosthetic for human intuition rather than as an independent generator of novel theory.

Of the many examples of AI-enabled advances in science, one could ask *what* exactly is being achieved: Is it better and more-useful models or predictors of natural phenomena than was possible otherwise? Is it the scale, scope, and speed of AI, along with the ability to learn from and explore vast spaces, without which the advances would have been practically infeasible or taken unfathomably long? Is it truly novel discoveries or breakthroughs on "grand challenge" problems? Or is it also the new scientific understanding we gain as a result? The answer clearly varies across the examples highlighted in this volume and beyond, with the majority (for now?) exhibiting the first couple of characteristics. Perhaps more examples of the latter kind will emerge, as more scientists use AI on more long-standing or new problems and in new ways (more on this later), or as AI's capabilities advance further.

How can science advance AI to further advance science?

In "Language Is Not All You Need," Joshua Tenenbaum critiques the orthodoxy that general intelligence emerges solely from scaling LLMs. He contrasts the "jaggedness" of current AI – which achieves superhuman results in some complex problems yet fails at elementary reasoning – with the robust, general intelligence of biological organisms. Positing that cognition evolutionarily precedes language, Tenenbaum argues that the field must transcend "scaling up" and emulate the biology of "growing up" (with echoes of Turing and Solomonoff). His synthesis of LLMs with Bayesian cognitive science and probabilistic programming aims to achieve the causal inference and generalization necessary for science. Going in the

other direction, he argues that such AI approaches may also "start to answer some of the fundamental questions about how our minds work and where our intelligence comes from." On the limits of language, it is worth noting Antonio Orvieto's critique in "Are Current AI Systems Unlocking Knowledge Discovery in Genomics?" that LLMs struggle with "state tracking" over the extremely long sequences found in genomics. Orvieto advocates for abandoning text-centric paradigms in favor of neural networks designed to achieve Turing-complete reasoning on biological data.[39] In her essay "Knowledge-Centric AI for Scientific Discovery," Carla Gomes contends that the prevailing data-centric paradigm of LLMs struggles to generalize beyond training distributions and adhere to inviolable scientific principles. By embedding scientific constraints directly into deep-learning architectures, she highlights novel advances in fields like materials science and ecology, including some that elude human intuition.

Progressing toward "Building an AI Polymath," Shirley Ho tackles a different challenge in which specialized AI models excel in specific domains yet fail to grasp nature's interconnectedness. Invoking Da Vinci and Newton, she is developing foundation models capable of reasoning across domains and scales, from quantum mechanics to fluid dynamics and astrophysics (see also Stella Offner's, Gege Wen's, and Sara Beery's perspectives on scales in their essays). Ho uses universal tokenization to put heterogeneous data archetypes into a unified latent space and uses symbolic distillation to translate neural computations into human-readable physical laws. Anima Anandkumar redirects our gaze toward the continuous nature of physical reality in her essay "How Do We Build AI to Push the Frontiers of Scientific Discovery?" She uses neural operators to go beyond discretization to learn mappings between continuous function spaces, allowing AI to internalize physical laws at multiple scales. Her stance offers a counterpoint to universal tokenization, asserting that the resolution of scientific discovery must be continuous and infinite.

In Surya Ganguli's "Toward a Science of Intelligence: Unifying Physics, Neuroscience & AI" and Maria Spiropulu and Hartmut Neven's "Quantum + AI = Quantum AI," the authors contend that further progress in AI requires a reexamination of its physical and biological substrates. Like Tenenbaum, Ganguli heeds the "taunting sirens" of biology, that human brains outperform AI by orders of magnitude in energy and data efficiency. He observes that unlike digital systems that require energy-intensive, reliable bit-flips, biological computation thrives on slow, unreliable, stochastic, and even sloppy intermediate steps, and suggests that this may constitute a parsimonious design feature that could benefit AI development, rather than a bug. And in a synthesis of neuroscience and condensed matter physics, he envisions neuromorphic architectures that aid discovery broadly and may even help elucidate consciousness. Spiropulu and Neven look to the fundamental nature of reality – building on foundations laid by the winners of the 2025 Nobel Prize in Physics and recent progress in quantum computing, in part enabled

by AI – and posit that the trajectory of AI inexorably leads to "Quantum AI." For example, they leverage quantum entanglement to develop out-of-time-order correlators (OTOCs) for elucidating molecular dynamics with novel scientific results.[40] Together, these essays suggest that for AI to be more capable for discovery, we will need architectures that reflect the fundamental nature of reality itself.[41]

Beyond this volume, many researchers are continuing to investigate the foundations of deep learning and its architectures, while others are building general assistive tools for scientists, many employing *in silico* agentic approaches, including literature-spanning multi-agent brainstorming, evolutionary mutations and tournaments, harnesses of various kinds, and neurosymbolic and surrogate modeling hybrids.[42] Yet others are focused on more foundational barriers to scientific discovery. Among the most formidable barriers is what Judea Pearl called the "ladder of causation."[43] Current deep learning is mostly on the first rung: *association*. Scientific discovery requires climbing to the rungs of *intervention* and *counterfactuals* (or imagining). One notable approach is Bernhard Schölkopf and his colleagues' causal representation learning, which merges the feature-learning of deep neural networks with the rigor of causal inference, an approach being applied, for example, to disentangle the genetic drivers of disease from environmental confounders.[44] But perhaps the hardest aspect of doing science (and for current AI systems) is coming up with novel conjectures and theories. Physicist Max Tegmark has argued that a model that predicts planetary orbits with 99.9 percent accuracy but cannot derive Kepler's laws is a tool for engineering, not physics. He and colleagues have been developing AI Feynman, a recursive symbolic regression algorithm to rediscover physical laws from data. In a long-standing tradition of testing AI via rediscovery of known science, in 2020, AI Feynman rediscovered one hundred equations from the Feynman Lectures on Physics.[45] Such directions suggest the possibility of novel theories and conjectures in a human-interpretable form.

What are the implications of AI-enabled science?

The integration of AI into the scientific enterprise represents more than enablement of discovery; it is also arguably a fundamental reconstitution of the social science of discovery – its applications and impacts on society, its institutions and people, its epistemic foundations and methods.[46] Most of the contributions to this volume engage these issues to varying degrees, some directly.

Eric Topol and Kelly Chibale focus on health care. In his essay "The Future of AI-Facilitated Medicine," Topol envisions a medical renaissance in which AI improves diagnostic precision and timing (that is, early detection). He points to clinical trials confirming, for example, that AI enhances breast cancer detection and that retinal scans can predict cardiovascular and neurodegenerative risks like Alzheimer's and Parkinson's disease in asymptomatic patients. Through cli-

nicians' "liberation from the keyboard" and delegation of "data clerk" functions to AI, Topol sees a future of revitalized doctor-patient relationships, yet notes the empathy paradox that AI poses. And just as there are developments toward AI co-scientists, so too toward AI "co-clinicians" as real-time multimodal assistive tools for care teams.[47] In "The Role of AI in Drug Discovery in Africa," Chibale argues that AI offers a unique mechanism to transition to innovation self-reliance. He advocates leveraging AI to address neglected diseases and Africa's genetic diversity, often not included in global datasets, and using AI-driven pharmacogenomics to rectify long-standing inequities in drug efficacy. He warns, however, that without local data, infrastructure, and access, the AI era risks exacerbating the health disparities it promises to resolve.[48]

In "Field Theory : AI as Social Science Question, Object & Tool," Alondra Nelson widens the volume's aperture. Invoking Max Weber and W. E. B. Du Bois, she frames AI not merely as an instrument but as a "social artifact" that reorganizes work, the production of knowledge, and the categories of social existence and meaning. She argues that AI poses urgent questions for the social sciences and demands a dual approach : as a methodological asset for useful, even beneficial, social inquiry and, crucially, as an object of study itself, including its effects on society. Recently, computational social scientist James Evans and colleagues framed LLMs as "cultural and social technologies" similar to the written word, economic markets, and state bureaucracies in that they do more than make information accessible ; they generate "lossy but useful representations" of unmanageably large and complex human-generated data and allow it to be reorganized, transformed, and restructured in ways that create opportunities and challenges for society and for discovery.[49] Returning to Nelson, she warns that without a robust social science of AI, we risk sociotechnical and sociological failures, including "narrowing of thought" and supplanting human judgment – issues M. J. Crockett also notes in their personal brief, "Making Automation Work for Social Scientists." Further, in their 2024 *Nature* paper with Lisa Messeri, Crockett spotlights the sociological failure of scientists overestimating their comprehension because an LLM provided them a plausible, fluent, or confident output. They also argue that the use of LLMs impacts other aspects of scientists' work – such as reading, reviewing, and writing – risking the creation of science monocultures.[50]

Arguing "Physics Is Different," Tess Smidt elucidates the "culture and craft" of doing physics, which she describes as not a monolith but a patchwork of distinct data cultures, infrastructures, and workflows. Contrasting the faintness of gravitational waves, the industrial-scale validation in particle physics, and the artisanal data of condensed matter physics, she argues that AI-enabled science remains inextricably bound to communal norms of discovery. Smidt also reminds us of the mismatches apparent in physics, that instruments (like LIGO or synchrotrons) take decades to build, while AI advances seemingly monthly, creating temporal friction.

In her essay "Thinking & Doing Science in the Age of AI," Alison Noble contends that AI impacts the scientific method and its workflows, transforming the scientist's role from creator to rigorous challenger of the AI's output.[51] She distinguishes between the doing of science – increasingly AI-assisted and automated – and the thinking of science, identifying peril where cognitive offloading erodes the "habit of thinking." Together, these perspectives offer a vital sociological counterweight to the technological views in other essays.

On the social science of discovery, two recent perspectives suggest other cautions and possibilities. Analyzing over forty million papers in the natural sciences, Qianyue Hao and colleagues reveal that while AI-enabled science expands researcher impact – tripling publications and quintupling citations compared to others – it narrows the collective aperture of inquiry, mostly because current AI capabilities direct research toward data-rich and established epistemologies over exploration of novel and data-sparse frontiers. Widening the aperture of inquiry requires "AI systems that expand not only cognitive capacity but also sensory and experimental capacity."[52] Elsewhere, Sendhil Mullainathan and Ashesh Rambachan argue that algorithms will reorganize science, especially the "patchwork sciences" – such as economics, medicine, and psychology – that do not converge toward a single "reigning champion" theory and rely on an arsenal of partial, often contradictory, context-dependent theories. They envision a "new factory floor" on which AI formalizes previously ineffable processes and revolutionizes discovery by learning which theories perform best and codifying the implicit judgment experts use to deploy their theoretical arsenal.[53]

Mario Krenn and Heather Champion pose ten questions for the "Philosophy of Autonomous Science." They posit that the (inevitable?) transition from AI tools to artificial scientists presents a profound philosophical challenge: translating epistemic traits – curiosity, surprise, creativity, taste, as well as notions of beauty or elegance – into computable, nonanthropocentric objectives. By confronting the "alien" nature of some AI discoveries and the interpretability challenges they entail, they call for a reevaluation of scientific understanding itself, questioning if human comprehension remains necessary for scientific progress. They propose moving beyond the use of AI as a tool – favored by some in this volume, along with the epistemic hierarchy that implies – toward the automation of epistemic norms. The transition is not unlike Aspuru-Guzik's trajectory from the era of "cyborg" scientists – human-machine hybrids that enable and extend human intent – to a future of "android" AI scientists capable of generating hypotheses and, with physical embodiment, executing experiments. Such visions are dually exhilarating and disquieting and, according to Krenn and Champion, may require humans to employ "Teacher AIs" to translate concepts beyond our comprehension, lest we find ourselves like the scientists in Ted Chiang's "Catching Crumbs from the Table," who are reduced to hermeneutics, merely interpreting the output of metahuman AI, with humanity

losing its grasp on the fundamental workings of the world. This could bring its own anxieties in that for many scientists, the pursuit of science is fueled by epistemic emotions: curiosity, awe, surprise, and the sheer "joy of insight" and understanding. If AI recursively self-improves to the point where humans are demoted to students of its discoveries, the future of discovery could face a crisis not of capability and useful progress but of human motivation and meaning.

In "After Science" – an essay in *Science* that would sit well alongside those in this section – James Evans and philosopher Eamon Duede suggest that while there has already been a considerable amount of "AI-infused science" that has led to advances, these have been bounded by human understanding and traditional methods of justification and evaluation. But as AI becomes more capable and novelty and surprise emerge, there could be advances beyond our comprehension or ability to explain. This would be a significant shift in which AI efficacy exists independently of human legibility, thus challenging our classical notions of scientific understanding.[54] One could further question the classical scientific method and its typical starting points: For example, in "Against Theory-Motivated Experimentation," computational cognitive scientist Marina Dubova and colleagues find that agents employing "gold-standard" or theory-driven scientific strategies – such as carefully premeditated experiments designed explicitly to falsify a dominant theory, confirm existing knowledge, or resolve a disagreement between two competing theories – create an "illusion of success," but perform worse than agents pursuing random and undirected exploration toward novel and useful theories.[55] This suggests that the more significant value of AI systems for scientific discovery may lie not in their ability to flawlessly execute standard, theory-driven, incremental hypothesis testing, however useful, but in their capacity for high-variance, unconstrained learning from and exploration of vast data, parameter spaces, and domains that human intuition (and theories) would naturally ignore for any number of reasons (prevailing orthodoxies and reigning theories, expertise or discipline blind spots, practical infeasibilities) that could lead to novel discoveries. Given how these introductory notes opened, such shifts and futures lead some to wonder: Who gets the Nobel Prize?[56]

The volume suggests other significant shifts in the future of discovery. For instance, the classic scientific method privileges empirical observation as the locus of discovery, with computation serving as a *post hoc* analytical tool. But many examples in biology and materials science suggest an ontological inversion whereby *in silico* approaches become the locus of discovery and the physical laboratory is demoted to a verification mechanism. In this way, the bottleneck for discovery shifts from the generation of ideas to the verification of ideas.[57] Here's another: Classically, the scientific method has leaned on reductionism, the ability to derive macroscopic phenomena from fundamental microscopic laws. However, phenomena at the mesoscopic scale – such as the dynamics of a cell or the behavior of turbulent fluids – suggest the possibility that discovery may no longer require solving con-

stitutive equations (like Schrödinger's or the Navier-Stokes) but rather navigating high-dimensional latent representations that predict emergent behaviors directly from data, as LeCun and others suggest. And perhaps most daring for its implications for discovery is Hassabis's contention that if neural networks can successfully reverse-engineer and simulate stable structures of nature, from protein folding to planetary dynamics, then perhaps "information is the most fundamental unit of physics, more so than matter and energy." This would suggest that physical laws (at least the ones we have) are not immutable axioms but emergent properties of a learnable informational process. If the universe is isomorphic to a Turing machine or a neural network, then the scientific method and discovery shift from observing a static material reality to decoding a dynamic, computational one.

Perhaps one should not view such shifts as a clarion call to abandon science, as it has been, and is, productively done in ways that advance our understanding of the world. Instead, such shifts – albeit, arguably of Copernican proportions – offer ways in which the scientific enterprise could be transformed in its scope, speed, and methods, and reimagined to significantly advance discovery.

Throughout the volume, there is an underlying rumble suggesting that the future of discovery in the age of AI calls not only for deeper foundations but for a reconceptualization of how scientists are educated, including overcoming the dichotomies between theoretical, computational, and experimental science; connecting and drawing from across disciplines; having a sophisticated understanding of state-of-the-art AI capabilities (and shortcomings) and an ability to engage in new workflows and rigorously evaluate outputs; and becoming framers of questions and architects of inquiry, imbued with a sociological curiosity and imagination regarding their tools, able to critique and contribute to advances in both directions at the intersection of AI and science. The "personal briefs" by early-career scientists provide a glimpse of such future scientists. While each offers a unique perspective, they go beyond the exuberance surrounding LLMs and advocate for science-informed AI. Yet their optimism is tempered by the "frictions" of their experience and the need to preserve epistemic agency. And finally, in "When AI Meets Art at the Scales of Science," Refik Anadol uses machine learning to render the architecture of collective scientific knowledge tangibly visible, translating vast, multiscalar datasets from neuroscience, molecular biology, and astrophysics into immersive sculptures.

Where does this leave us? I hope full of enthusiasm to explore the possibilities and challenges as well as the tensions and provocations that the volume raises for the future of discovery, and to explore aspects not covered in this volume. For not only does the future of discovery matter for science, it matters for humanity's progress and prosperity.

In his 2025 Nobel Prize lecture, economic historian Joel Mokyr makes a distinction between *propositional knowledge* (science) and *prescriptive knowledge* (technology), argu-

ing that new technology drives waves of scientific advances, which in turn drive the development of new technologies, in a "positive feedback loop that can keep growing with no end in sight" – *and* it does so in ways that have historically enhanced welfare and propelled humanity's progress.[58] The interaction of AI (prescriptive knowledge) and science (propositional knowledge) demonstrates this positive feedback loop, with two characteristics that together could further expand the possibilities for progress: AI is among a few technologies with a strong claim to being both a general-purpose technology and an invention of methods of invention.[59] In this regard, perhaps what lies ahead ought to be thought of as a combination of the Industrial Revolution (whose general purpose technologies enabled economic progress) and the Enlightenment (for the scientific method, instruments for discovery, and its "Copernican shifts"). This not only places AI in the mainstream of technological advances that have driven both profound change, and, importantly, humanity's progress throughout history, but does so with novel and concurrent potency.

While these characteristics along with the advances highlighted and proposed in this volume project a future for discovery and societal progress full of possibilities, none of it is guaranteed or automatic – or without complexities, disruptions, and risks. On one hand, it will require developing AI that is even more capable of advancing science and enabling many more scientists in more places to participate at the expanding intersection of AI and science, not only to advance discovery in both directions but to harness it to tackle humanity's greatest *current* and future opportunities and challenges, and benefit people everywhere – an aspect not always guaranteed, even when discovery and progress occur, as history has shown. On the other hand, it will be critically important to tackle apace the complexities of how AI reconfigures the sociological, structural and institutional, and epistemological aspects of discovery, and to take seriously the risks – including safety and ethics – that have always been at the intersection of scientific discovery and powerful technologies, taking into account the unique characteristics of AI. Both aspects will require investment in science and collaboration across academic, private-sector, and government laboratories. In addition, the issues noted at the outset that are giving rise to unease in this new era of science (and of AI generally) must be tackled so as not to impair the advancement of science, its applications, and our trust in it. All this will need active shaping by the scientific community and others outside it with focus not only on what could go wrong or be lost, but also what could go right and be gained.[60] Echoing an idea attributed to another Nobel laureate, physicist Dennis Gabor, our question *What is the future of discovery?* is not a question of prediction but of design.[61]

AUTHOR'S NOTE

I am grateful to the American Academy for the opportunity to conceive this *Dædalus* volume and to bring together perspectives from some of the scientists at the frontier of AI & Science. And I am grateful to the contributors. On a theme as broad as this, there are undoubtedly many more scientists and topics that are absent; for that I take responsibility. I found inspiration and learned from the contributors to this volume, from others through our conversations and collaborations–including Jeff Dean, Stuart Russell, Yoshua Bengio, Blaise Agüera y Arcas, Fei-Fei Li, Mira Murati, Dario Amodei, Dan Huttenlocher, Eric Schmidt, Erik Brynjolfsson, Nigel Shadbolt, Wendy Hall, Meghan O'Sullivan, Jennifer Doudna, Mike Spence, Michel Devoret, and Gillian Tett (especially for the access to Turing's papers and notes at King's College, Cambridge)–and from the work of many scientists, especially those cited in the endnotes. I would like to thank my colleagues at Google Research and Google DeepMind for our work together at the intersection of AI & Science, as well as colleagues at AI2050 and Stanford's AI Index. I am especially grateful to Lizzie Dorfman (a scientist colleague whose research is also cited in this volume), who was a valuable thought-partner throughout. And as with the 2022 volume on AI & Society, this volume on AI & Science would not have come together without generous collaboration and partnership with the Academy's talented editorial team of Phyllis Bendell, Editor of *Dædalus*, who brought her experience as guide and editor and enthusiasm from the very beginning to the completion of this effort, and Peter Walton and Maya Robinson, who were creative and expert copyeditors for all the essays in this volume. These introductory notes benefited from valuable feedback from Lizzie Dorfman, Blaise Agüera y Arcas, Kerry McHugh, Kent Walker, Sarah Manyika, and Julian Manyika, but they should not be held responsible for any errors or opinions herein. The views expressed are mine and do not necessarily reflect the views of Google or Alphabet.

ABOUT THE AUTHOR

James Manyika, a Member of the American Academy since 2019, is SVP at Google-Alphabet. He is also President for Research, Labs, Tech & Society, focusing on Google and Google DeepMind's foundational and applied research and innovations in AI, computing, and science. Most recently, he served as Vice Chair of the U.S. National AI Advisory Committee established by Congress to advise the President on AI (2022–2025) and as Co-Chair of the UN Secretary General's High-level Advisory Body on AI (2023–2024). He is an inaugural Distinguished Fellow of Stanford's Human-Centered AI Institute and a Distinguished Fellow in Ethics in AI at Oxford, where he is also an appointed Visiting Professor. He is the guest editor of the Spring 2022 issue of *Dædalus* on "AI & Society."

ENDNOTES

[1] See Demis Hassabis, "Accelerating Scientific Discovery with AI," Nobel Prize lecture, Aula Magna, Stockholm University, December 8, 2024, https://www.nobelprize.org/prizes/chemistry/2024/hassabis/lecture; and David Baker, "De Novo Protein Design," Nobel

Prize lecture, Aula Magna, Stockholm University, December 8, 2024, https://www.nobelprize.org/prizes/chemistry/2024/baker/lecture.

[2] For the role that AI "hallucinations" played in Baker's work, see William J. Broad, "How Hallucinatory AI Helps Science Dream Up Big Breakthroughs," *The New York Times*, December 23, 2024, https://www.nytimes.com/2024/12/23/science/ai-hallucinations-science.html. AI diffusion models such as those used by Baker are a physics-inspired deep-learning approach co-introduced in 2015 by one of the contributors to this volume, Surya Ganguli. See Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli, "Deep Unsupervised Learning Using Nonequilibrium Thermodynamics," arXiv (2015), https://arxiv.org/abs/1503.03585.

[3] For examples of AI-enabled advances beyond those in this volume, see the annual *Artificial Intelligence Index Reports* (2017–2026) by the AI Index initiative (of which I am on the steering committee) at the Stanford Institute for Human-Centered AI, available at https://hai.stanford.edu/ai-index/2026-ai-index-report.

[4] See Rick Stevens, Valerie Taylor, Jeff Nichols, et al., *AI for Science: Report on the Department of Energy (DOE) Town Halls on Artificial Intelligence (AI) for Science* (U.S. Department of Energy Office of Scientific and Technical Information, 2019), https://www.osti.gov/biblio/1604756; The Royal Society, *Science in the Age of AI: How Artificial Intelligence Is Changing the Nature and Method of Scientific Research* (The Royal Society, 2024); National Academies of Sciences, Engineering, and Medicine, *Foundation Models for Scientific Discovery and Innovation: Opportunities Across the Department of Energy and the Scientific Enterprise* (National Academies Press, 2023), https://doi.org/10.17226/29292; National Science Foundation, "National Artificial Intelligence Research Institutes" (consisting of, at time of publication, twenty-nine AI institutes across multiple science disciplines) (accessed March 11, 2026); and U.S. Department of Energy, "Genesis Mission," https://genesis.energy.gov (announced in late 2025, "A National Mission to Accelerate Science through Artificial Intelligence" involving seventeen national laboratories, academia, and the private sector).

[5] See Hanchen Wang, Tianfan Fu, Yuanqi Du, et al., "Scientific Discovery in the Age of Artificial Intelligence," *Nature* 620 (7972) (2023), https://doi.org/10.1038/s41586-023-06221-2.

[6] See Philip Ball, "Is AI Leading to a Reproducibility Crisis in Science?" *Nature* 624 (2023): 22–24, https://doi.org/10.1038/d41586-023-03817-6; Arvind Narayanan and Sayash Kapoor, "Why an Overreliance on AI-Driven Modelling is Bad for Science," *Nature* 640 (2025): 312–314, https://doi.org/10.1038/d41586-025-01067-2; and Editorial, "AI Scientists Are Changing Research–Institutions, Funders and Publishers Must Respond," *Nature* 651 (2026): 853–854, https://doi.org/10.1038/d41586-026-00934-w. On safety, see Yoshua Bengio, Stephen Clare, Carina Prunkl, et al., *International AI Safety Report 2026* (UK Government, 2026), https://internationalaisafetyreport.org/sites/default/files/2026-02/international-ai-safety-report-2026_1.pdf; and National Academies of Sciences, Engineering, and Medicine, *The Age of AI in the Life Sciences: Benefits and Biosecurity Considerations* (National Academies Press, 2025), https://doi.org/10.17226/28868.

[7] AI Index, *Artificial Intelligence Index Reports* (2022–2026).

[8] See the reports of the UN Secretary General's High-level Advisory Body on AI, which I cochaired: Advisory Body on Artificial Intelligence, *Interim Report: Governing AI for Humanity* (United Nations, 2023), https://www.un.org/sites/un2.un.org/files/un_ai_advisory_body_governing_ai_for_humanity_interim_report.pdf; and Advisory Body on Artificial Intelligence, *Governing AI for Humanity* (United Nations, 2024), https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf.

[9] See John McCarthy, Marvin L. Minsky, Nathaniel Rochester, and Claude E. Shannon, "A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence," August 31, 1955, http://raysolomonoff.com/dartmouth/boxa/dart564props.pdf. The aims of the project read today like the pursuit of artificial general intelligence: "An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves." The participants included scientists from neuroscience, cognitive science, cognitive psychology, mathematics, computer science, and information engineering. See my lecture on the seventy-fifth anniversary of the Turing Test at Kings College, Cambridge, "The Next Turing Tests," October 15, 2025, https://www.kingselab.org/all-events/james-manyika. See also Eddy Keming Chen, Mikhail Belkin, Leon Bergen, and David Danks, "Does AI Already Have Human-Level Intelligence? The Evidence Is Clear," *Nature* 650 (2026): 36–40, https://doi.org/10.1038/d41586-026-00285-6; and Blaise Agüera y Arcas and James Manyika, "AI Is Evolving–And Changing Our Understanding of Intelligence," *Noema*, April 8, 2025, https://www.noemamag.com/ai-is-evolving-and-changing-our-understanding-of-intelligence.

[10] Marvin Minsky, *The Emotion Machine: Commonsense Thinking, Artificial Intelligence, and the Future of the Human Mind* (Simon & Schuster, 2007).

[11] See President's Information Technology Advisory Committee, *Computational Science: Ensuring America's Competitiveness* (Executive Office of the President of the United States, 2005), https://www.nitrd.gov/pubs/pitac/pitac_report_computational-science_2005.pdf; and W. Brian Arthur, "Algorithms and the Shift in Modern Science," Beijer Discussion Paper Series No. 269 (Beijer Institute of Ecological Economics, Royal Swedish Academy of Sciences, 2020), https://beijer.kva.se/wp-content/uploads/2020/03/Disc269_Arthur_2020.pdf. See "Theory and Observation in Science," Stanford Encyclopedia of Philosophy.

[12] An early illustration of this was Stanislaw Ulam and John von Neumann's work at Los Alamos in the 1940s that resulted in the Monte Carlo method and its use in computational physics and in Monte Carlo tree search (MCTS), formalized in 2006 and popularized by AlphaGo in 2016. See Iulia Georgescu, "The Forgotten Pioneers of Computational Physics," *Physics World*, November 11, 2025, https://physicsworld.com/a/the-forgotten-pioneers-of-computational-physics; and David Silver, Aja Huang, Chris J. Maddison, et al., "Mastering the Game of Go with Deep Neural Networks and Tree Search," *Nature* 529 (7587) (2016): 484–489, https://doi.org/10.1038/nature16961. See also our recent combination of MCTS and a foundation model to solve a class of general science problems: Eser Aygün, Anastasiya Belyaeva, Gheorghe Comanici, et al., "An AI System to Help Scientists Write Expert-Level Empirical Software," *Nature* (2026), https://doi.org/10.1038/s41586-026-10658-6.

[13] See the chapter "The Pre-history of Computing" in Blaise Agüera y Arcas, *What Is Intelligence? Lessons from AI about Evolution, Computing, and Minds* (MIT Press, 2025).

[14] See Alan M. Turing, "On Computable Numbers, with an Application to the Entscheidungsproblem," *Proceedings of the London Mathematical Society* S2-42 (1) (1937): 230–265, https://doi.org/10.1112/plms/s2-42.1.230; Alan Turing, "'Intelligent Machinery, A Heretical Theory,' a Lecture Given to '51 Society' at Manchester," AMT/B/4, The Turing Digital Archive, https://turingarchive.kings.cam.ac.uk/publications-lectures-and-talks-amtb/amt-b-4; Alan M. Turing, "The Chemical Basis of Morphogenesis," *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 237 (641) (1952): 37–72; John von Neumann, *The Computer and the Brain* (Yale University Press, 1958); and John von Neumann, *Theory of Self-Reproducing Automata*, ed. Arthur W. Burks (University of Illinois Press, 1966).

[15] Several prominent physicists share this view, including John Wheeler, Leonard Susskind, Paul Davies, and 2022 Nobel laureate Anton Zeilinger.

[16] See Stephen Wolfram, "Cellular Automaton Fluids 1 : Basic Theory," *Journal of Statistical Physics* 45 (3/4) (1986). See also the chapters "Life as Computation" and "Artificial Life" in Agüera y Arcas, *What Is Intelligence?*

[17] See Allen Newell and Herbert A. Simon, *The Logic Theory Machine : A Complex Information Processing System* (RAND Corporation, 1956). Pamela McCorduck, in "Artificial Intelligence : An Aperçu," *Dædalus* 117 (1) (1988), tells how when the Logic Theorist proved the *Principia* theorems, Bertrand Russell (author of *The Principia Mathematica*) was delighted, but "*The Journal of Symbolic Logic* declined to publish an article – coauthored by the Logic Theorist – that described the proof," anticipating some of today's questions of authorship.

[18] Margaret O. Dayhoff, Richard V. Eck, Marie A. Chang, and Minnie R. Sochard, *Atlas of Proteins Sequence and Structure* (The National Biomedical Research Foundation, 1965). See also Robert K. Lindsay, Bruce G. Buchanan, Edward A. Feigenbaum, and Joshua Lederberg, "DENDRAL : A Case Study of the First Expert System for Scientific Hypothesis Formation," *Artificial Intelligence* 61 (1993) : 209–261. For more of the history of the period from 1950–1975 and its personalities, see Pamela McCorduck, *Machines Who Think*, 2nd ed. (Routledge, 2004).

[19] Barto and Sutton's work, along with Geoffrey Hinton's, influenced my early research and first publication. See Robert K. Appiah, Jean H. Daigle, James M. Manyika, and Themuso Makhurane, "Modelling and Training of Artificial Neural Networks," *African Journal of Science and Technology Series B, Science* 6 (1) (1992) ; and most recently with Seijin Kobayashi, Yanick Schimpf, Maximilian Schlegel, et al., "Emergent Temporal Abstractions in Autoregressive Models Enable Hierarchical Reinforcement Learning," arXiv (2025), https ://arxiv.org/pdf/2512.20605. In 2025, I had the honor of presenting Andrew Barto and Rich Sutton with the Turing Award for their work on reinforcement learning at the Association of Computing Machinery (ACM) awards ceremony.

[20] Hubert L. Dreyfus and Stuart E. Dreyfus, "Making a Mind Versus Modeling the Brain : Artificial Intelligence Back at a Branchpoint," *Dædalus* 117 (1) (Winter 1988) : 15–44, https ://www.amacad.org/sites/default/files/daedalus/downloads/Daedalus_Wi98_Artificial-Intelligence.pdf.

[21] Jeffrey Dean, "A Golden Decade of Deep Learning : Computing Systems & Applications," *Dædalus* 151 (2) (Spring 2022), https ://www.amacad.org/publication/daedalus/golden-decade-deep-learning-computing-systems-applications. Foundational advances in deep learning include several by pioneers in the field whose papers are considered influential and foundational classics of this period and referred to in the field in shorthand as : AlexNet (2012), Word2Vec (2013), DQN (2013/15), VAEs (2014), GANs (2014), Adam (2014), ResNet (2015), AlphaGo (2016), Transformer (2017), BERT (2018), and GPT-3 (2020).

[22] Kevin Scott, "I Do Not Think It Means What You Think It Means : Artificial Intelligence, Cognitive Work & Scale," *Dædalus* 151 (2) (Spring 2022), https ://www.amacad.org/publication/daedalus/i-do-not-think-it-means-what-you-think-it-means-artificial-intelligence-cognitive-work-scale. See Ray J. Solomonoff, "A Formal Theory of Inductive Inference, Part I," *Information and Control* 7 (1) (1964), https ://raysolomonoff.com/publications/1964pt1.pdf ; and Ray J. Solomonoff, "A Formal Theory of Inductive Inference, Part II," *Information and Control* 7 (2) (1964), https ://raysolomonoff.com/publications/1964pt2.pdf. See also Universal Artificial Intelligence (AIXI) in Peter Sunehag and Marcus Hutter, "Principles of Solomonoff Induction and AIXI," in *Algorithmic Probability and Friends. Bayesian*

Prediction and Artificial Intelligence, ed. David L. Dowe (Springer Nature Link, 2011). Both Solomonoff induction and AIXI have informed our work; see Alexander Meulemans, Rajai Nasser, Maciej Wołczyk, et al., "Embedded Universal Predictive Intelligence: A Coherent Framework for Multi-Agent Learning," arXiv (2025), https://arxiv.org/pdf/2511.22226.

[23] For other examples of genomics advances in this "golden decade," see the Telomere-to-Telomere Consortium, which decoded the final 8 percent of the human genome to produce the first complete human reference genome, using Google Research's DeepVariant to polish the assembled sequence: Sergey Nurk, Sergey Koren, Arang Rhie, et al., "The Complete Sequence of a Human Genome," *Science* 376 (6588) (2022), https://doi.org/10.1126/science.abj6987. See also the use of DeepVariant and DeepConsensus by a consortium including my colleagues to create the first draft human pangenome: Wen-Wei Liao, Mobin Asri, Jana Ebler, et al., "A Draft Human Pangenome Reference," *Nature* 617 (7960) (2023), https://doi.org/10.1038/s41586-023-05896-x.

[24] See Jared Kaplan, Sam McCandlish, Tom Henighan, et al., "Scaling Laws for Neural Language Models," arXiv (2020), https://doi.org/10.48550/arXiv.2001.08361; Dario Amodei, "Machines of Loving Grace," October 2024, https://www.darioamodei.com/essay/machines-of-loving-grace; and Dario Amodei, "The Adolescence of Technology," January 2026, https://www.darioamodei.com/essay/the-adolescence-of-technology. See also Lee Sharkey, Bilal Chughtai, Joshua Batson, et al., "Open Problems in Mechanistic Interpretability," arXiv (2025), https://arxiv.org/pdf/2501.16496.

[25] "Ilya Sutskever Declares 'Game Over for Scaling Laws'–AI Enters New Era," AIM Network, YouTube video, November 26, 2025, https://www.youtube.com/watch?v=UROxY4ceHlo. On System 1 and 2 thinking, see Daniel Kahneman, *Thinking Fast and Slow* (Farrar, Straus and Giroux, 2011).

[26] See Fei-Fei Li, "Form Words to Worlds: Spatial Intelligence is AI's Next Frontier," November, 2025, https://drfeifei.substack.com/p/from-words-to-worlds-spatial-intelligence.

[27] See Usman Anwar, Abulhair Saparov, Javier Rando, et al., "Foundational Challenges in Assuring Alignment and Safety of Large Language Models," arXiv (2024), https://doi.org/10.48550/arXiv.2404.09932. See also Yoshua Bengio, Michael Cohen, Damiano Fornasiere, et al., "Superintelligent Agents Pose Catastrophic Risks: Can Scientist AI Offer a Safer Path?" arXiv (2025), https://doi.org/10.48550/arXiv.2502.15657.

[28] Stuart Russell and Peter Norvig, *Artificial Intelligence: A Modern Approach*, 4th ed. (Pearson, 2020). See also Stuart Russell, "If We Succeed," *Dædalus* 151 (2) (Spring 2022): 43–57, https://www.amacad.org/publication/daedalus/if-we-succeed; and Stuart J. Russell, *Human Compatible: Artificial Intelligence and the Problem of Control* (Viking, 2019).

[29] Rich Sutton, "The Bitter Lesson," Incomplete Ideas, 2019, http://www.incompleteideas.net/IncIdeas/BitterLesson.html.

[30] See I. J. Good's paper "Speculations Concerning the First Ultra Intelligent Machine" (1964), in which he discusses the possibility of recursive self-improvement and the notion that such an AI would lead to an "intelligence explosion" and be "the last invention man need ever make." He also argued that such a machine would "give the human race a good chance of surviving indefinitely" but also acknowledged "the opposite possibility, that the human race will become redundant." See also Jack Clark, "AI Systems Are about to Start Building Themselves: The First Step Towards Recursive Self-Improvement," Import AI 455, May 4, 2026, https://importai.substack.com/p/import-ai-455-automating-ai-research.

[31] Another notable advance in inorganic materials is MatterGen by Microsoft Research. See Claudio Zeni, Robert Pinsler, Daniel Zügner, et al., "A Generative Model for Inorganic Materials Design," *Nature* 639 (2025): 624–632, https://doi.org/10.1038/s41586-025-08628-5.

[32] See Veronika Eyring, William D. Collins, Pierre Gentine, et al., "Pushing the Frontiers in Climate Modelling and Analysis with Machine Learning," *Nature Climate Change* 14 (2024), https://doi.org/10.1038/s41558-024-02095-y. For examples of Earth foundation models, see Cristian Bodnar, Wessel P. Bruinsma, Ana Lucic, et al., "A Foundation Model for the Earth System," *Nature* 641 (8065), https://doi.org/10.1038/s41586-025-09005-y (Aurora by Microsoft Research); and Daniel Klocke, Claudia Frauen, Jan Frederik, et al., "Computing the Full Earth System at 1 km Resolution," arXiv (2025), https://doi.org/10.48550/arXiv.2511.02021 (ICON Earth system by NVIDIA/MaxPlank Institute). See also work with my colleagues: Aaron Bell, Amit Aides, Amr Helmy, et al., "Earth AI: Unlocking Geospatial Insights with Foundation Models and Cross-Modal Reasoning," arXiv (2026), https://arxiv.org/pdf/2510.18318 (Earth AI); and Christopher F. Brown, Michal R. Kazmierski, Valerie J. Pasquarella, et al., "AlphaEarth Foundations: An Embedding Field Model for Accurate and Efficient Global Mapping from Sparse Label Data," arXiv (2025), https://arxiv.org/abs/2507.22291v2 (AlphaEarth Foundations).

[33] See also Nicholas Stern (author of the influential 2006 *Stern Review on the Economics of Climate Change*) and his colleagues in Nicholas Stern, Mattia Romani, Roberta Pierfederici, et al., "Green and Intelligent: The Role of AI in the Climate Transition," *npj Climate Action* 4 (2025), https://doi.org/10.1038/s44168-025-00252-3.

[34] Matthew Hutson, "AI Hunts for the Next Big Thing in Physics," *IEEE Spectrum*, February 3, 2026, https://spectrum.ieee.org/particle-physics-ai. See also the recent consensus study report of the National Academies, which notes the current and future use of AI for discovery: National Academies of Sciences, Engineering, and Medicine, *Elementary Particle Physics: The Higgs and Beyond* (National Academies Press, 2025), https://doi.org/10.17226/28839.

[35] In "Can AI Solve Science?" Stephen Wolfram argues that while AI allows us to tackle the plethora of computationally reducible problems, the existence of problems that are computationally irreducible means that to fully advance science we must "combine the strengths of AI and of the formal computational paradigm." See Stephen Wolfram, "Can AI Solve Science?" March 5, 2024, https://writings.stephenwolfram.com/2024/03/can-ai-solve-science.

[36] From Turing's lecture to the London Mathematical Society, February 20, 1947. See also *Nature* Editorial, "Why AI Cannot Do Good Science Without Humans," *Nature*, May 19, 2026, https://doi.org/10.1038/d41586-026-01551-3. For an example, see Syed Asad Rizvi, Daniel Levine, Aakash Patel, et al., "Scaling Large Language Models for Next Generation Single-Cell Analysis," bioRxiv (2025), https://doi.org/10.1101/2025.04.14.648850, by Yale and Google Research.

[37] For example, see the use of AlphaTensor to break a fifty-year record in matrix multiplication using tensor decomposition as a reinforcement learning game in Alhussein Fawzi, Matej Balog, Aja Huang, et al., "Discovering Faster Matrix Multiplication Algorithms with Reinforcement Learning," *Nature* 610 (7930) (2022), https://doi.org/10.1038/s41586-022-05172-4. In hashing and sorting, see AlphaDev in Daniel J. Mankowitz, Andrea Michi, Anton Zhernov, et al., "Faster Sorting Algorithms Discovered Using Deep Reinforcement Learning," *Nature* 618 (7964) (2023), https://doi.org/10.1038/s41586-023-06004-9. In NP-Hard combinatorics, see FunSearch, which uses LLMs as

intelligent mutation engines embedded within an evolutionary algorithm, in Bernardino Romera-Paredes, Mohammadamin Barekatain, Alexander Novikov, et al., "Mathematical Discoveries from Program Search with Large Language Models," *Nature* 625 (2024), https://doi.org/10.1038/s41586-023-06924-6.

[38] See Bogdan Georgiev, Javier Gómez-Serrano, Terence Tao, and Adam Zsolt Wagner, "Mathematical Exploration and Discovery at Scale," arXiv (2025), https://doi.org/10.48550/arXiv.2511.02864; and Romera-Paredes, Barekatain, Novikov, et al., "Mathematical Discoveries from Program Search with Large Language Models." See also Benjamin Skuse, "AI Is Acing Math Exams Faster than Scientists Write Them," *IEEE Spectrum*, February 25, 2026, https://spectrum.ieee.org/ai-math-benchmarks.

[39] For other recent genetics advances, see Nauman Javed, Thomas Weingarten, Arijit Sehanobish, et al., "A Multi-Modal Transformer for Cell Type-Agnostic Regulatory Predictions," *Cell Genomics* 5 (2) (2025); Žiga Avsec, Natasha Latysheva, Jun Cheng, et al., "Advancing Regulatory Variant Effect Prediction with AlphaGenome," *Nature* 649 (8099) (2026), https://doi.org/10.1038/s41586-025-10014-0; and Garyk Brixi, Matthew G. Durrant, Jerome Ku, et al., "Genome Modelling and Design across All Domains of Life with Evo 2," *Nature* 652 (8112) (2026), https://doi.org/10.1038/s41586-026-10176-5.

[40] See papers coauthored with my colleagues and our collaborators on quantum and related science advances: Google Quantum AI and Collaborators, "Quantum Error Correction Below the Surface Code Threshold," *Nature* 638 (8052) (2025), https://doi.org/10.1038/s41586-024-08449-y; Google Quantum AI and Collaborators, "Observation of Constructive Interference at the Edge of Quantum Ergodicity," *Nature* 646 (8086) (2025), https://doi.org/10.1038/s41586-025-09526-6; C. Zhang, R. G. Cortiñas, A. H. Karamlou, et al., "Quantum Computation of Molecular Geometry via Many-Body Nuclear Spin Echoes," arXiv (2025), https://doi.org/10.48550/arXiv.2510.19550; and Charina Chou, James Manyika, and Hartmut Neven, "The Race to Lead the Quantum Future," *Foreign Affairs*, January/February 2025, https://www.foreignaffairs.com/united-states/race-lead-quantum-future-chou-manyika-neven.

[41] See also Mohammad Ghazi Vakili, Christoph Gorgulla, Jamie Snider, et al., "Quantum-Computing-Enhanced Algorithm Unveils Potential KRAS Inhibitors," *Nature Biotechnology* 43 (12) (2025), https://doi.org/10.1038/s41587-024-02526-3.

[42] For examples of such tools, see AI Scientist in Chris Lu, Cong Lu, Robert Tjarko Lange, et al., "Towards End-to-End Automation of AI Research" *Nature* 651 (2026): 914–919, https://doi.org/10.1038/s41586-026-10265-5; CORAL in Ao Qu, Han Zheng, Zijian Zhou, et al., "CORAL: Towards Autonomous Multi-Agent Evolution for Open-Ended Discovery," arXiv (2026), https://doi.org/10.48550/arXiv.2604.01658; and SAGA in Yuanqi Du, Botao Yu, Tianyu Liu, et al., "Accelerating Scientific Discovery with Autonomous Goal-Evolving Agents," arXiv (2026), https://doi.org/10.48550/arXiv.2512.21782. See also our work in Juraj Gottweis, Wei-Hung Weng, Alexander Daryin, et al., "Accelerating Scientific Discovery with Co-Scientist," *Nature* (2026), https://doi.org/10.1038/s41586-026-10644-y.

[43] Judea Pearl and Dana Mackenzie, *The Book of Why: The New Science of Cause and Effect* (Basic Books, 2020).

[44] See Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, et al., "Towards Causal Representation Learning," https://arxiv.org/abs/2102.11107; and Goutham Rajendran, Simon Buchholz, Bryon Aragam, et al., "From Causal to Concept-Based Representation

Learning," in *Advances in Neural Information Processing Systems 37 (NeurIPS 2024)*, ed. Amir Globerson, Lester Mackey, Danielle Belgrave, et al. (Curran Associates, Inc., 2024).

[45] See Silviu-Marian Udrescu and Max Tegmark, "AI Feynman : A Physics-Inspired Method for Symbolic Regression," *Science Advances* 6 (16) (2020).

[46] See Sebastian Musslick, Laura K. Bartlett, Suyog H. Chandramouli, et al., "Automating the Practice of Science : Opportunities, Challenges, and Implications," *Proceedings of the National Academy of Sciences* 122 (5) (2025), https://doi.org/10.1073/pnas.2401238121. See also, from 1996, Gillies, *AI and the Scientific Method* ; and his 2022 update given AI progress, Donald Gillies, "Artificial Intelligence and Philosophy of Science from 1990s to 2020," in *Current Trends in Philosophy of Science : A Prospective for the Near Future*, ed. Wenceslao J. Gonzalez (Synthese Library, 2022).

[47] See recent results from field studies in breast cancer detection in collaborations between Imperial College London, the UK's National Health Service (NHS), and Google Research : Lucy M. Warren, Jenny Venton, Kenneth C. Young, et al., "Impact of Using Artificial Intelligence as a Second Reader in Breast Screening Including Arbitration," *Nature Cancer* 7 (2026), https://www.nature.com/articles/s43018-026-01128-z ; and Christopher J. Kelly, Marc Wilson, Lucy M. Warren, et al., "Diagnostic Accuracy, Fairness and Clinical Implementation of AI for Breast Cancer Screening : Results of Multicenter Retrospective and Prospective Technical Feasibility Studies," *Nature Cancer* 7 (2026), https://www.nature.com/articles/s43018-026-01127-0. On tools for clinicians, see Derek C. Angus, Rohan Khera, Tracy Lieu, et al., "AI, Health, and Health Care Today and Tomorrow : The JAMA Summit Report on Artificial Intelligence," *JAMA* 334 (18) (2025). See work with my colleagues : Khaled Saab, Chunjong Park, Tim Strother, et al., "Advancing Conversational Diagnostic AI with Multimodal Reasoning," *Nature Medicine* (2026), https://doi.org/10.1038/s41591-026-04371-0 ; and Elahe Vedadi, David Barrett, Natalie Harris, et al., "Towards Physician-Centered Oversight of Conversational Diagnostic AI," arXiv (2025), https://doi.org/10.48550/arXiv.2507.15743 (forthcoming in *Nature*). See also David Autor and James Manyika, "A Better Way to Think About AI," *The Atlantic*, August 24, 2025, https://www.theatlantic.com/technology/archive/2025/08/ai-job-loss-human-enhancement-google/683963.

[48] Georgia Channing and Avijit Ghosh, "AI for Scientific Discovery Is a Social Problem," *Patterns* 7 (3) (2026), https://doi.org/10.1016/j.patter.2026.101497.

[49] Henry Farrell, Alison Gopnik, Cosma Shalizi, and James Evans, "Large AI Models are Cultural and Social Technologies," *Science* 387 (6739) (2025), https://doi.org/10.1126/science.adt9819.

[50] See Lisa Messeri and M. J. Crockett, "Artificial Intelligence and Illusions of Understanding in Scientific Research," *Nature* 627 (8002) (2024), https://doi.org/10.1038/s41586-024-07146-0 ; and Nitya Thakkar, Mert Yuksekgonul, Jake Silberg, et al., "A Large-Scale Randomized Study of Large Language Model Feedback in Peer Review," *Nature Machine Intelligence* (2026), https://doi.org/10.1038/s42256-026-01188-x.

[51] In David P. Woodruff, Vincent Cohen-Addad, Lalit Jain, et al., "Accelerating Scientific Research with Gemini : Case Studies and Common Techniques," arXiv (2026), https://doi.org/10.48550/arXiv.2602.03837, we tackle, with academic collaborators, several open problems in computer science, physics, optimization, and economics with some novel results and highlight new workflows and the necessity of human orchestration. See also Don Knuth, "Claude's Cycles" (2026), https://www-cs-faculty.stanford.edu/~knuth/papers/claude-cycles.pdf, in which Knuth (Turing Award winner, considered the founding father of the rigorous analysis of algorithms) and his colleagues

use Claude to solve an open combinatorial problem about directed Hamiltonian cycles that he had been stuck on.

[52] Qianyue Hao, Fengli Xu, Yong Li, and James Evans, "Artificial Intelligence Tools Expand Scientists' Impact but Contract Science's Focus," *Nature* 649 (8099) (2026), https://doi.org/10.1038/s41586-025-09922-y. See also Sabina Leonelli and Alexander Mussgnug, "Convenience AI," PhilSci Archive (2025), https://philsci-archive.pitt.edu/24891.

[53] Sendhil Mullainathan and Ashesh Rambachan, "Science in the Age of Algorithms," in *The Economics of Transformative AI*, ed. Ajay K. Agrawal, Erik Brynjolfsson, and Anton Korinek (forthcoming with University of Chicago Press). See also Nicola Jones, "Half of Social-Science Studies Fail Replication Test in Years-Long Project," *Nature*, April 1, 2026, https://doi.org/10.1038/d41586-026-00955-5. For an example of AI use in the historical sciences, see Donal Khosrowi and Finola Finn, "Can Generative AI Produce Novel Evidence?" *Philosophy of Science* 92 (5) (2025), https://doi.org/10.1017/psa.2025.10123.

[54] James Evans and Eamon Duede, "After Science," *Science* 390 (6774) (2025), https://doi.org/10.1126/science.aec7650.

[55] Marina Dubova, Arseny Moskvichev, and Kevin Zollman, "Against Theory-Motivated Experimentation in Science," MetaArXiv (2025), https://doi.org/10.31222/osf.io/ysv2u_v2.

[56] See Dashun Wang, "Prizes Must Recognize Machine Contributions to Discovery," *Nature* 646 (2025), https://doi.org/10.1038/d41586-025-03217-y.

[57] A defining characteristic of late-twentieth-century philosophy of science is the dissolution of the strict dichotomy between theory and observation, with many embracing the "theory-ladenness" of observation–the realization that empirical results are inherently the result of prior theoretical commitments and values, and the limitations of the instruments and signal-processing algorithms used to produce them. See "Theory and Observation in Science," Stanford Encyclopedia of Philosophy, rev. January 12, 2026, https://plato.stanford.edu/entries/science-theory-observation. See also discussions on the philosophical status of simulations and *in silico* experiments in "Computer Simulations in Science," Stanford Encyclopedia of Philosophy, rev. February 19, 2026, https://plato.stanford.edu/entries/simulations-science.

[58] See Joel Mokyr, "The Past and Future of Innovation: Can Progress Be Sustained?" Nobel Prize lecture, Aula Magna, Stockholm University, December 8, 2025, https://www.nobelprize.org/prizes/economic-sciences/2025/mokyr/lecture.

[59] See Martin Neil Baily, David M. Byrne, Aidan T. Kane, and Paul E. Soto, "Generative AI at the Crossroads: Light Bulb, Dynamo, or Microscope?" (Brookings Institution, 2025), https://www.brookings.edu/wp-content/uploads/08/2025/baily-byrne-kane-soto_generative-ai-crossroads_2025-09-05_FINAL-WEB.pdf; and Nicholas Crafts, "Artificial Intelligence as a General-Purpose Technology: An Historical Perspective," *Oxford Review of Economic Policy* 37 (3) (2021), https://doi.org/10.1093/oxrep/grab012.

[60] On the need for academic and private sector collaboration for AI for science, see James Manyika, "Bidirectional Collaboration," part of Marcia McNutt, "For a More Competitive U.S. Research Enterprise, the Work Begins Now," *Issues in Science and Technology* 41 (1) (2024), https://issues.org/state-of-the-science-mcnutt. See also James Manyika, "Getting AI Right: A 2050 Thought Experiment," in *The Digitalist Papers: Artificial Intelligence and Democracy in America* (Stanford Digital Economy Lab, 2024), https://www.digitalistpapers.com/essays/getting-ai-right.

[61] Dennis Gabor, *Inventing the Future* (Secker & Warburg, 1963).