MAI

SCIENCE SERIES

What do we mean when we say AI for science?

At MAI, we believe that AI will accelerate the rate of scientific breakthroughs and usher in a new age of discovery and transformative technology. But to make bold claims like that, we must also present a practical vision. In this piece, we break down the scientific process into stages and examine where AI is genuinely accelerating progress, what challenges could slow it down, and what the future of AI-accelerated science looks like.

October 2026

AI is transforming scientific research at an exhilarating pace that has no historical precedent. In just a small handful of years, it's been reported that AI systems have solved the structure of pretty much every known protein,[1] designed entire bacteriophage genomes,[2] and discovered brand new materials.[3] AI is driving progress in long-pondered mathematical conjectures,[4] predicting weather with weekslong time horizons,[5] and building digital twins of fruit fly brains.[6] Huge national initiatives like the U.S. Genesis Mission[7] show that AI for science is moving from individual demonstrations to an entire national research strategy across academic, federal, and private labs.

These headlines are astounding, and we certainly believe that AI will drive a cascade of new scientific breakthroughs. But to construct a vision that prepares for this future of AI-accelerated science, we must also be cautious of overpromises and present the pragmatic view. This report is our attempt at that, by exploring the future of scientific discovery through the different dimensions of what "AI for science" looks like. In this piece, we first decompose the scientific research process and examine how AI tools are catalyzing transformations in the different stages of discovery. AI is already conducting literature searches; generating hypotheses; and speeding up simulations, experiments, and analyses—all across different fields, skills, and scales. But the ecosystem of science may be slower to adapt and may pose bottlenecks that limit wider impact. We discuss potential pitfalls related to

---

[1] Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021).

[2] King, S. H. et al. Generative design of bacteriophages with genome language models. Science 393, eaec2657 (2026).

[3] Merchant, A. et al. Scaling deep learning for materials discovery. Nature 624, 80–85 (2023).

[4] OpenAI. On the Navier-Stokes Millennium Prize Problem. (2026). Available at: https://openai.com/index/navier-stokes-solution/

[5] Bodnar, C. et al. A foundation model for the Earth system. Nature 641, 1180–1187 (2025).

[6] Vaxenburg, R. et al. Whole-body physics simulation of fruit fly locomotion. Nature 643, 1312-1320 (2025).

[7] The White House. Launching the Genesis Mission. Executive Order 14363 (2025). Available at: https://www.whitehouse.gov/presidential-actions/2025/11/launching-the-genesis-mission/

ground truth data, trust, communication, and measurement of scientific progress, and we present a set of policy recommendations for consideration.

AI is changing science, there's no doubt about that. But in this rapidly evolving space of AI capabilities, discoveries, and promises, we present an examination of what becomes abundant, and what becomes scarce in the future of science.

Abundant: Scientific iteration

Scientific breakthroughs can often seem like they emerge by serendipity—a paradigm shift in understanding that comes in a Eureka moment—but science is more appropriately described as a process of trial and iteration. Scientific research is a cycle of collecting prior knowledge, generating hypotheses, creating and testing models, designing experiments, taking measurements, and interpreting results. For any topic of inquiry, projects shift between these stages, not always sequentially and spending more time in some stages than others—and sometimes going straight back to the drawing board. But in general, the more we try, the more we know. Remarkably, AI is massively accelerating each of these traditional steps, allowing for rapid iteration and creating wide swaths of new discovery space.

However, it should be stated here early on that the scientific method looks different within every discipline, even if the shape may be roughly consistent—and even in this work, we take a perspective more aligned with the experimental physical sciences rather than theoretical sciences or social sciences. When considering the different dimensions of research, there is considerable heterogeneity between theoretical frameworks, data availability, verifiability, and the complexity of physical experimentation. And AI and AI tools will look different in each field as well—in some AI may mean large language models, in others small specialized models, and in others comprehensive world models. Promising results in one field should not be immediately taken as evidence of generalizability across multiple scientific fields, and certain research questions will be more amenable to AI discovery than others even within a field.

COLLECTING PRIOR KNOWLEDGE

AI's highly developed language capabilities can be readily applied to synthesizing and understanding prior scientific knowledge, combining disparate sources like textbooks, research articles, industry reports, blogs, conferences, and more, and including multi-disciplinary and multi-lingual analysis. Deep literature reviews which may have previously taken two years can now be done in as little as a month.[8] AI can also connect to existing datasets and databases to build new analyses more easily. Researchers can now gain a thorough understanding of the field and open questions faster, making it easier to advance to the frontier of knowledge and to begin experimenting.

GENERATING HYPOTHESES

Hypotheses build on prior knowledge to generate testable claims about how the world works, often asking why or how a pattern emerges. Often—though certainly not always—hypotheses arise combinatorially: a mishmash of ideas and techniques from other research, applied to new questions with new tools or new data.[9] In this sense, AI tools can scan the literature for relevant information, combine information from different fields, and generate novel hypotheses. Multi-agent frameworks can also judge generated hypotheses by creating a tournament-style Elo score for hypotheses, comparing generated hypotheses against one another to construct a ranking of promising ideas.[10] The top-ranked ideas can then be further evolved, combined, and ultimately tested.

However, earlier research in the field of Natural Language Processing suggested that while AI-generated hypotheses may look more novel, exciting, and effective than expert human-generated hypotheses at first glance, after 100 hours of investigation, they are downrated more.[11] And using tournament-style techniques to judge hypotheses against one another can select for surface-level novelty rather

---

[8] Anthropic. Claude Science, an AI workbench for scientists, is now available. Anthropic (2026). Available at: https://www.anthropic.com/news/claude-science-ai-workbench

[9] Simonton, D. K. Scientific Creativity: Discovery and Invention as Combinatorial. Front. Psychol. 12, 721104 (2021).

[10] Gottweis, J. et al. Accelerating scientific discovery with Co-Scientist. Nature 655, 487-496 (2026).

[11] Si, C., Hashimoto, T. & Yang, D. The Ideation-Execution Gap: Execution Outcomes of LLM-Generated versus Human Research Ideas. Preprint at https://doi.org/10.48550/arXiv.2506.20803 (2025).

than deeper insight. So while recent results in biomedical science and computational biology offer strong opportunity that multi-agent frameworks are able to generate hypotheses and experiments on par with experts,[12] their generalizability across different fields is not yet explored.

MODELS

In the scientific sense, a model refers to a simplified picture of reality that can arise both as a hypothesis or as an experimental probe of reality. In AI, a model refers to the large bundle of weights and parameters that similarly reflect a simplified picture of reality, but maybe in a slightly different sense. But AI models play a significant role in constructing, testing, or analyzing our scientific models, especially from an angle of computational experimentation. These can be split into predictive and generative AI models, explained below.

Predictive models are AI models trained on large amounts of experimental data or physics-based calculations so they can then predict the outcomes of an experiment or calculation without undertaking the whole process. This can accelerate the process massively, save orders of magnitude of resources, or even make previously intractable problems, tractable. For example, AlphaFold predicted protein structures for over 200 million sequences, from an original 170,000 experimentally determined protein structures aggregated over decades through x-ray diffraction and other measurements.[13] GNoME predicted the structural energy of 2.2 million crystalline atomic arrangements and suggested that 380,000 of them may be stable materials—against a dataset of just 20,000 experimentally known crystalline materials and 28,000 additional calculated materials.[14]

Predictive AI models can also dramatically speed up or make possible simulations of complex physical dynamics that for previous numerical techniques required intensive supercomputing resources. Without AI, this would become intractable as the system got more complex or the calculations more precise, for example for high-precision weather forecasting. But Microsoft's Aurora AI system—trained on

---

[12] Gottweis, J. et al. Accelerating scientific discovery with Co-Scientist. Nature 655, 487-496 (2026)

[13] Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583-589 (2021).

[14] Merchant, A. et al. Scaling deep learning for materials discovery. Nature 624, 80-85 (2023).

decades of Earth systems data—has demonstrated accurate weather, atmospheric, and oceanic predictions at a fraction of the compute cost and with higher resolutions and longer lead-times, including predicting severe storms like tropical cyclones.[15]

Predictive models solve the problem, "From this initial condition, what is the outcome?" whereas generative design models flip that: "For this target outcome, what do I start with?" For example, Microsoft Research's MatterGen model generates structures of inorganic crystalline materials that have a desired electronic or mechanical property.[16] Models have also emerged for protein design[17] to specify binding targets, shape, stability, and more; mRNA design[18] to specify translation efficiency or suitability for therapeutic applications; and DNA design to specify functional constraints.[19]

Critical to existing generative design models is that the system can be discretized into building blocks with quantifiable characteristics—like tokens in language models, think atoms or electrons described by their spatial location, or how genetic materials are described by their nucleotide sequence. This discretization allows the same family of generative techniques that we see in language to be applied to molecular, materials, and biological domains.

PHYSICAL EXPERIMENTS AND MEASUREMENTS

Predictive and generative AI models must be paired with real-world experimentation and validation, such as by synthesizing the predicted materials or genetic sequences. Scientific equipment for synthesis, fabrication, characterization, and analysis is being increasingly automated by AI—which then also outputs a continuous stream of structured data for training future models. While currently autonomous labs may look mostly like robotic arms, in the near future there will be end-to-end autonomous systems which can do closed-loop

---

[15] Bodnar, C. et al. A foundation model for the Earth system. Nature 641, 1180–1187 (2025).

[16] Zeni, C. et al. A generative model for inorganic materials design. Nature 639, 624–632 (2025).

[17] Watson, J. L. et al. De novo design of protein structure and function with RFdiffusion. Nature 620, 1089–1100 (2023).

[18] Zhang, H. et al. Deep generative models design mRNA sequences with enhanced translational capacity and stability. Science 390, eadr8470 (2025).

[19] Brixi, G. et al. Genome modelling and design across all domains of life with Evo 2. Nature 652, 1349–1361 (2026).

experimentation across a range of conditions, able to select experimentation variables, characterize outputs, and build learnings without human intervention.[20]

SOFTWARE TOOLS AND ANALYSIS

AI coding tools are already transforming computational and empirical sciences, bringing low- and no-code software capabilities to every scientist. AI coding tools also connect with existing scientific databases like for bioinformatics and existing software packages for domain-specific analysis. Now even fields that were not typically code-heavy can quickly and easily construct simulations, data analysis pipelines, classification and labeling algorithms, data visualizations, and more, which can increase the scale or breadth of experiments that can be conducted. This also closes traditional divides between computational scientists and non-computational scientists, with a faster iteration loop between computational and physical world experimentation.

Abundant: The AI scientist

The previous section described AI tools which follow the current paradigm of science, complementing and accelerating existing workflows. As AI systems improve further—with the natural world uniting with natural language such as through vision-language-action models and robotics—they will operate with increasing autonomy and efficiency through every stage outlined previously. In the next wave, combined "AI scientist" systems will tackle open-ended science problems end-to-end by reasoning through new hypotheses, constructing experiments to validate them, and autonomously analyzing results and proposing iterations.[21] As they improve in capability and access to physical machinery, AI scientists will begin to autonomously bridge gaps between fields and scales to connect fundamental, microscopic theories to system-level understanding and transformative technological insights. And this will all happen at scale, working

---

[20] Szymanski, N. J. et al. An autonomous laboratory for the accelerated synthesis of inorganic materials. Nature 624, 86–91 (2023).

[21] Gottweis, J. et al. Accelerating scientific discovery with Co-Scientist. Nature 655, 487–496 (2026).
Ghareeb, A. E. et al. A multi-agent system for automating scientific discovery. Nature 655, 497–505 (2026).

24/7 and massively in parallel—an entirely new structure of how science is done that will accelerate the pace of ideas, discoveries, and knowledge generation.

Abundant: Multi-disciplinary, multi-scale science

Until now, our scientific understanding has been in part limited to how scientists communicate new findings to one another, which can be limited by language access[22] and field-specific jargon and siloed into the choice of department, publication, or conference that scientists self-organize into.

AI tools are bridging these divides, and aggregating knowledge through cross-domain analyses is likely to yield new perspectives and patterns about the natural world. As a fun historical example of a cross-scale insight (pre-AI), take allometric scaling laws: that even with the incredible complexity of living organisms, some biological traits like metabolic rate, brain size, heart rate, growth rate, genome length, and more, follow a simple power law scaling with body mass.[23] So from elephants down to the respiratory complexes in our cells, the same mathematical relationship between mass and metabolism holds over an astounding 27 orders of magnitude.[24] Many physical laws remain true across disciplines, with different fields each having their unique lens on the phenomena. It seems possible that AI will uncover more of such relationships, simply due to its ability to take a macroscopic lens across fields.

Another example of opportunity comes from the history and economics of science. Ben Jones's "Burden of Knowledge" found that scientists' age at first innovation (measured through patent data) has been increasing, suggesting the informal theory that as knowledge requirements deepen, innovators pursue longer educational periods and specialize to a greater degree in order to get to the

[22] There's a joke in physics: if you ever discover a new phenomenon in condensed matter physics, first check the footnotes of Soviet journals from the 1960s.

[23] West, G. B. & Brown, J. H. Life's Universal Scaling Laws. Physics Today 57, 36–42 (2004).

[24] Ibid.

frontier.[25] But it is well known that analytical techniques developed in one field may also be applicable to disparate fields—this has been a long-standing goal of interdisciplinary sciences. But training any individual to be an expert in many domains becomes increasingly difficult, as the burden of knowledge demands increasing specialization and longer training. AI tools are already making way for cross-disciplinary opportunities: OpenAI's disproof of the Erdős unit distance conjecture applied a technique from algebraic number theory to a geometry problem.[26]

Scarce: Ground truth data & data deserts

While the previous sections explored a new abundance of knowledge and discovery, even in this richness there may be a scarcity of certain types of data, especially verifiable, ground truth data. A tantalizing aspect of AI for science is the exploration of brand new physical parameter spaces: like wavelengths smaller than even our best microscopes, or more complex interactions and dynamics than we've measured before, or hotter temperatures than is reachable by traditional experiment. It will be tempting to use AI and data-driven tools to extrapolate discoveries into these new physical frontiers. But just because it's simulatable by AI doesn't mean it's verifiable or practically useful. AI discoveries must also be paired with innovations in instrumentation that can maintain ground truth data from the real world to continuously validate and improve AI models. This includes things like weather sensors, geological sensors, quantum sensors, advanced microscopes, and materials characterization techniques. Advancing scientific data collection can also feed into developing new AI models, but for this flywheel to unfold we must not let AI substitute for continued investments in fundamental sciences, methods development, and instrumentation engineering.

It should also be carefully considered how scientific AI models will inherit the biases and blindspots of their training datasets and methods. For example, current

[25] Jones, B. F. The Burden of Knowledge and the "Death of the Renaissance Man": Is Innovation Getting Harder? Review of Economic Studies 76, 283–317 (2009).

[26] OpenAI. An OpenAI model has disproved a central conjecture in discrete geometry. OpenAI (2026); available at https://openai.com/index/model-disproves-discrete-geometry-conjecture/

materials prediction models are commonly trained on the outputs of density functional theory (DFT) calculations, which solve a computationally-intensive system of quantum mechanical equations that scale exponentially with atomic size. Supercomputers have been buzzing away on DFT calculations for decades, producing a treasure trove of data. But DFT itself is an approximation, and is known to be insufficient for certain materials like superconductors, magnetic materials, and layered materials. Also, DFT is cheapest and most systematic for crystalline materials and smaller molecules, hence existing datasets are comprised overwhelmingly of crystalline materials and not amorphous materials, polymers, or disordered materials which are commonly found in real-world applications. So, an overreliance on DFT calculations for materials prediction could miss a whole swath of promising materials that aren't crystalline.

This exacerbates issues with data deserts: regions of parameter space where only limited, fragmented, or low-quality data exists and therefore predictions in this space may be poorly grounded. Data deserts may exist for any number of reasons, such as low historical collection, lower funding, operational difficulty, or technical feasibility. As a concrete example, many parts of the globe like parts of Africa, South Asia, and Latin America lack the dense network of weather and ocean sensors that climate AI models are built off of, which impedes their accuracy and early-warning capabilities.[27] Issues with data coverage are further exacerbated for things like medical data, where data availability is tied to data governance, and also defy simple classification. And in many cases of complex, multidimensional data, it may not even be known yet what the full parameter space comprises. So while AI may help extend coverage into traditional data deserts, it is still important to be mindful of downstream impacts.

Scarce: Trust & attention

Science depends not only on producing new claims, but also on deciding which claims deserve attention and trust. Peer review has typically been a gold standard

---

[27] Jain, A. AI-Powered Weather Forecasting: The New Frontier of Climate Intelligence. UChicago Sustainability Dialogue (2025). Available at: https://sustainabilitydialogue.uchicago.edu/news/ai-powered-weather-forecasting-the-new-frontier-of-climate-intelligence/

for the acceptance of claims. However, as research topics have become more complex and increasingly specialized, peer review has become more of a signal of trust—more like a governance mechanism that affirms that the submitted work is authentic and presents a good faith effort—rather than a full scrutiny or reproduction of claims. Earlier researchers have noted these flaws of peer review, including human biases and a presumption of trust against critical errors and fraud.[28]

Researchers are now observing an increase in LLM-supported drafting across the spectrum of quality.[29] This increase in volume—and also as AI enables increasing amounts of cross-disciplinary, highly-specialized, and convincing-sounding research—means it will be increasingly difficult to find peer reviewers with the requisite expertise as well as the time and incentives to review.[30]

In this respect, more than half of reviewers are already using AI for peer review,[31] though without clear guidance to ensure AI tools augment the review rather than undermine the validity or introduce new biases. The scientific publishing industry must adapt to these AI transformations in order to maintain trust in scientific reporting and the integrity of the scientific process. It should become standard for journal editors to develop and use AI review tools to cross-check related works and to check technical grounding, statistical power, and for signatures of data manipulation. Peer review (by human peers) should remain paramount for critiquing experimental design and interpretation of results. Submissions and reviews should be made public and be both human- and machine-readable for independent auditing.

Moreover, the use of journal articles as the primary means of conveying findings will likely become insufficient. All results, including null and negative results and incomplete datasets, should be reported in a machine-readable format and with sufficient metadata so that the analysis can be independently reconstructed. Data

---

[28] Smith, R. Peer Review: A Flawed Process at the Heart of Science and Journals. J R Soc Med 99, 178–182 (2006).

[29] Kusumegi, K. et al. Scientific production in the era of large language models. Science 390, 1240–1243 (2025).

[30] Aczel, B. et al. The present and future of peer review: Ideas, interventions, and evidence. Proc. Natl. Acad. Sci. U.S.A. 122, e2401232121 (2025).

[31] Naddaf, M. More than half of researchers now use AI for peer review — often against guidance. Nature 649, 273–274 (2026).

should be in standardized formats such that data from different authors and sources can be easily combined into a broad scientific corpora. And as generative AI models make it both easier to access and execute strong, interesting research that gets rigorously cross-checked, so too will it be easier to produce low-quality research that may look convincing without sufficient scrutiny. Increasingly, research will have to be accompanied by signifiers of trust, confidence, rigor, and reproducibility in the research. In modern science, these signifiers have typically been related to peer review, individual or institutional prestige, experience, and reported methods. However, this has had its limitations as well: for example, we have seen this have a homogenizing effect on science training.[32] With AI, it's not yet clear what these signifiers will be, and whether they may be equalizers or unequalizers compared to the current status quo. For example, earlier research has shown that institutional prestige affects how quickly or far new scientific findings spread, even when controlled for quality of findings.[33] AI may enable teams with lower resources and unaffiliated with a typical prestigious institution to conduct groundbreaking research and for it to have large impact; but also, a flood of AI-generated research may also lead people to over-index on prestige as the only available (though not necessary accurate) signal of quality. So while new discoveries may become abundant, the trust and attention required to understand and discern these may be in short supply.

Scarce: Measurement of scientific progress

As AI reduces the cost of producing hypotheses, analyses, code, and manuscripts, it becomes more important to understand and measure the impact of such artifacts and whether they are materially advancing new discoveries and catalyzing new technologies. This has strong implications as well for how science is funded, which currently is through a combination of federal, industry, and philanthropic streams, and on how to select the most promising ideas to seed and pursue. It is

---

[32] Nowogrodzki, J. Most US professors are trained at same few elite universities. Nature 609, 887-887 (2022).

[33] Morgan, A. C., Economou, D. J., Way, S. F. & Clauset, A. Prestige drives epistemic inequality in the diffusion of scientific ideas. EPJ Data Sci. 7, 40 (2018).

already difficult to measure scientific progress. Metrics like papers, citations, patents, grant dollars, and journal prestige tend to measure activity more so than new knowledge or impact, especially when impacts may take decades to develop and build on many innovations simultaneously. Additionally, it may be increasingly important to measure the creativity or diversity of new ideas in science—however this may be defined—in order to support or rebut arguments that AI will have a homogenizing or self-defeating effect on scientific discovery.

The scientific community should continue to track the diversity and flow of ideas, such as through proxies like patents, collaboration maps, and citation networks. Additionally, scientific discoveries should be tracked long-term for their impact on technology development and applications.

Policy recommendations

Treat scientific data and provenance as infrastructure

Increasing usage of AI models and autonomous scientific machinery will create exponential amounts of new data. This is most likely a good thing. However, it comes with a responsibility to organize this data and to carefully track metadata and provenance. Scientific communities should create and maintain domain-specific interoperable standards for metadata and data availability, and where possible create standards for protocols involved. Instrument manufacturers and creators of scientific software should be incentivized to follow universal schema, such as with sustained support of open-source software and standards. Data management should be invested in like infrastructure—built to be shared, long-lived, and foundational for future advances.

Build new institutions for communication, verification, and replication

It appears likely that traditional journals and peer review will be insufficient for communicating the new pace of scientific discoveries as well as models and data. We should fund new efforts and experiments in research communications to understand what new signifiers for trust and attention will be, including for communicating beyond narrow scientific communities and to the general public,

including technologists, policymakers, and LLMs. Replication, robustness, and transferability checks should be funded adequately and pursued independently. The standard for trusted, verified discoveries will likely increase—and timelines and funding must be concomitantly updated as well.

Fund validation through to pilots

There is a known "valley of death" between ground-breaking scientific discoveries and technology that makes its way into people's lives. This is because as innovations move along this path, the funding needs, the incentives, and the skills involved also evolve. The ecosystems that work for initial discoveries are often not sufficient for developing it into technology at scale. One way to ease this is through funding continuity, shared validation facilities, and incentivized pilot demonstrations. This could look like shared infrastructure for small-batch manufacturing, process-engineering support for scientists, and milestone-based funding for flexible teams. Having this ecosystem in place would help demonstrate that promising AI-catalyzed scientific discoveries can be technologically feasible at least at a pilot scale, easing its way to larger-scale commercialization.

Avoid the concentration of capital and an erosion of small science

If scientific breakthroughs become increasingly linked with AI development—such as with more capable models, more compute-intensive models and agentic workflows, or bespoke models that rely on large amounts of specialized data—then this may also then be highly concentrated in industry R&D labs that have access to compute and autonomous labs. This would align research pursuits and the benefits of research breakthroughs with industry incentives, while eroding the critical roles that public and philanthropic science play. IP protections would then entrench research concentrations in private companies. Continued vertical integration of the entire science and technology stack—with no public obligation—poses a strong risk for widely accessible, societally transformative technological breakthroughs, especially where market incentives may be weak.

It would be a mistake not to invest in research infrastructure that enables smaller science efforts for small institutions and even for individuals. Previous research

has suggested that smaller teams are more likely to introduce disruptive directions, while larger teams are more likely to develop established ones.[34] While AI will certainly enable individuals and small teams to do more than they ever could before, another tension will play out: you can do more with less, but you need more to do more. The scientific enterprise must pursue a balance of these efforts and support institutions of all sizes.

Conclusion

AI for science is not just one thing, but rather comprises a tremendously exciting slate of advances and speedups to nearly every facet of the traditional scientific process, creating new models of the world and new modes of discovery. But in so far as science seeks to describe and understand the real world beyond silicon simulations, it will require real-world verification. The institutions around the scientific enterprise, including experimental facilities, data infrastructure, verification, and communication, must adapt rapidly to this new wave of scientific discovery. And if they can do that—with both science and scientific institutions transformed by AI—it appears more likely than not that we will witness a tremendous amount of AI-enabled discovery in the coming years such as in medicine, energy, chemistry, engineering, and more that will change the way we understand and interact with the world.

[34] Wu, L., Wang, D. & Evans, J. A. Large teams develop and small teams disrupt science and technology. Nature 566, 378–382 (2019).