

# Charting a Course for AI in Science

Rather than haphazardly using artificial intelligence for research, scientists and decisionmakers should take a deliberative approach.

Artificial intelligence is reshaping the scientific enterprise. From our perch at the Center for Security and Emerging Technology (CSET), a tech policy research organization at Georgetown University, we are witnessing this change through two lenses simultaneously. As researchers, we study AI, including large language models (LLMs) like those that power ChatGPT, and scientific models like AlphaFold; we also monitor the science and tech landscape, including trends in AI's development. Our ability to do this work relies in part on designing, testing, and implementing AI-enabled methods. AI is thus both our research subject and one of our research tools. Our work brings us face-to-face with the tensions that arise when figuring out what AI-enabled science can—and should—look like.

AI allows us to do work that previously required prohibitive amounts of time, money, and manual effort. For example, our Map of Science—an interactive visualization tool to explore trends in global science and technology—relies on LLMs to extract, clean, and curate metadata from hundreds of millions of scholarly publications. We also maintain LLM-enabled workflows to support our policy analysis, which includes translating, summarizing, and classifying documents ranging from government contracts and Chinese-language news articles to global AI governance policies. Where we previously spent months building data-

cleaning scripts and iteratively developing human annotation guides, we now spend only weeks or even days designing, testing, and refining prompts and validating outputs.

But this AI-enabled work only further highlights the value of our human researchers and their specialized knowledge. The functionality and utility of the Map of Science, for example, is made possible by a team that decides what data to include; designs and conducts robust review and validation of outputs; and determines how to deploy and document resulting systems. Given the many ways that data can be misunderstood, human insight and careful decisionmaking are required.

The tensions we are witnessing in our own work are mirrored across science right now: for individual researchers and teams, across fields, and for the enterprise as a whole. AI is breaking new ground, with applications ranging from biological research to deciphering the language of sperm whales, examining political attitudes, and measuring how narratives spread. But the push to use AI for science is exacerbated by factors beyond its transformative power. Resources are flowing to AI-driven research, even as funding for other research is under threat. Major science funders are prioritizing AI as an area for study, and the US government is signaling support for AI-enabled science through initiatives like the Genesis Mission, which aims to accelerate scientific breakthroughs through the use of AI. At the same time,

entrenched “publish or perish” incentives, narrow measures of scientific productivity, and growing demand to communicate work across platforms and audiences are rewarding speed and volume—making AI tools attractive regardless of their impact on rigor or understanding.

These forces are putting immense pressure on researchers to use AI. According to a 2025 survey of global researchers, 58% use AI tools in their work, up from 37% in 2024. Yet only 27% worldwide believe they have adequate training in those tools, and only 22% of US respondents think it will improve the quality of their work. Researchers and scientists are largely navigating these changes on their own, with current incentives pushing them to adopt AI through hurried learning of new and insufficiently understood tools, methods, and models.

Rather than integrating AI and LLMs into science in a haphazard and scattershot way, the scientific community needs a more deliberative approach that includes researchers and decisionmakers across the enterprise. The appetite for such discussions was made clear in a 2025 workshop in which CSET teamed with the Open Research Community Accelerator (ORCA) to convene experts in AI, open science, and metascience.

of AI use: from the need to update standards for scholarly attribution and licensing to the way autonomous AI labs are designed and resourced.

The potential for increased efficiency from AI use comes with trade-offs. Hurried, ad hoc incorporation of AI risks prioritizing speed over rigor, scale over understanding, and outputs over insight. In practice, such use often means outsourcing core research tasks such as literature review, data collection, analysis, and manuscript generation to AI. Handing over these basic scholarly responsibilities risks not gaining the familiarity and understanding required for the careful hypothesis generation, interpretative judgment, and deep engagement with relevant research that underpins scientific rigor and expertise. If not matched with updated structures and standards, such use of AI could entrench existing shortcomings of the research enterprise, while rewarding a narrow set of achievements, notably publication output. Uneven access to AI tools and infrastructure could deepen inequities between institutions, disciplines, and regions, further concentrating power and influence among a small subset of researchers with resources.

## The push to use AI for science is exacerbated by factors beyond its transformative power. Resources are flowing to AI-driven research, even as funding for other research is under threat.

Attendees shared their own use of AI and, in candid moments, wondered about the effects of outsourcing thinking and reasoning. They recounted AI use in their work—from proposal drafting and writing code to data annotation and cleaning—and highlighted issues surfaced by these examples. In one case, exploratory use of AI-generated participant consent documents for clinical trials raised accountability and accuracy concerns. Another discussion focused on agentic AI research assistants that demonstrated a lack of transparency and causal reasoning. And in a third example, employing LLMs for de-identifying clinical data for public use uncovered issues with model bias. Using AI in research also raised more practical issues, from increasing human review burden to being computationally expensive.

Participants wondered what would happen if total review and reasoning authority were delegated to AI; what happens when AI use is so pervasive it bypasses disclosure and evaluation; and whether people might end up skipping the thinking step altogether. Since that workshop, we’ve been in many discussions where scientists, administrators, funders, and others have shared examples and debated the implications

Careless and opaque use of AI could also erode trust and exacerbate public concerns around transparency, reproducibility, and validity in science. If researchers, policymakers, and the public cannot understand, audit, or confidently interpret how scientific claims are generated, the legitimacy of science as a shared societal enterprise is weakened. Many research tasks have no ground truth, or answer key, to evaluate AI outputs. Comparing LLM and human performance on research tasks can be undermined by insufficiently rigorous and transparent measurement of human capabilities. And given that AI systems can amplify preexisting human biases, distorted or inaccurate outputs may go undetected by researchers.

Because AI is altering all stages of the research process, holistic efforts are needed to actively steer its integration into science. At each step, the community should openly discuss the central question around AI-enabled science: what AI *can* do versus what it *should* do.

For example, AI can be used to shape science at the very first stages of the research process, from accelerating literature reviews with AI research assistants to generating

hypotheses and designing experiments with the help of AI scientists. Companies are adapting their LLMs to serve researchers, creating modules for peer-reviewed scientific content or connecting them to common software tools, and advertising such capabilities as “research partners” or “co-scientists.” While AI has been shown to generate new ideas and experimental designs, these successes are accompanied by cases of critical failures when attempting to validate claims and evaluate novel hypotheses. AI also has a propensity to generate exaggerated hypotheses and hallucinate citations. But beyond the ways that the technology fails, it raises other questions about its effect upon researchers: Reliance on AI for literature review and hypothesis generation will change how scientists and researchers learn from prior work and ideate new theories.

Similarly, although AI shows promise for conducting research, the benefits of automation may take a toll on researchers and the research environment. As AI systems are used for data collection and experimentation, they are enabling self-driving labs that optimize and conduct chemical and biological experiments. These tools are

these shortcomings when deriving conclusions; they must assume that every solved structure is likely “wrong” in some way and modify their conclusions accordingly. But the larger problem is that many AI systems generate predictions with limitations that are not always apparent and often without explaining how they arrived at them. As a tool to generate predictions that inform conclusions, AI must be coupled with the human scientists’ ability to contextualize and interpret.

Finally, AI is changing how research is vetted and communicated. These are likely the areas where current general-purpose LLMs can have the broadest impact, taking on tasks like summarizing manuscripts for social media posts or assisting research funders in checking for policy compliance or reviewing a scientist’s contributions. An important example is peer review. Long held as the gold standard for research dissemination, the peer review system is under strain, faced with increasing submissions (including an influx of fraudulent papers), overtaxed reviewers, and rising costs. To ease pressure points, publishers are turning to AI to handle aspects of the

## The potential for increased efficiency from AI use comes with trade-offs. Hurried, ad hoc incorporation of AI risks prioritizing speed over rigor, scale over understanding, and outputs over insight.

even enabling research with human subjects using “AI interviewers” that conduct interviews and develop questions in real time via a chat interface, at a scale and speed not possible with human interviewers. However, these efficiency gains come with serious concerns about ethical research conduct. Integrating AI into data collection and experimentation, especially with human subjects, will require updates to scientific oversight and standards across fields. These examples resurface the tension between the roles of human and machine, demanding reconsideration of the value of face-to-face engagement and field research, both for research subjects and as part of the experience of being a researcher.

AI is also helping researchers derive conclusions and interpret results. Protein structure prediction offers a widely known example. Determining a protein structure in the lab is an unpredictable process: Proteins are too small to look at directly, even under a microscope, so the method relies on measuring proxies under artificial conditions. AI models can predict structures rapidly and efficiently, but only at the cost of obscuring real-life messiness. The structural biologists who use such models are trained to account for

refereeing process. AIP Publishing, the American Physical Society, and IOP Publishing have piloted AI-driven citation checking, metadata validation, and ethics compliance. Meanwhile, Wiley, Elsevier, and other publishers are exploring ways to flag AI-generated papers and using AI to detect paper mills. But in a world where publication remains the key metric for career advancement, researchers sometimes lean on AI to supplement or even create their manuscripts. Incorporating AI into the review and production of scientific writing again brings up the tension of what is best done by humans and what could be done by AI.

At all of these stages, AI is changing the research process in ways that demand fresh thinking about scientific standards and processes. AI is also redefining what it means to be a researcher: what knowledge and skills are required, what ethical behavior entails, and what careers look like. It can redefine the size, shape, and scope of the scientific enterprise: how resources are allocated, how quality control operates, and how the system defines and values concepts like speed, rigor, and novelty. Perhaps most significantly, AI may redefine the relationship between science and society.

To meet this moment, stakeholders across the research and development ecosystem—including from academia, government, philanthropy, and industry—need sustained, shared forums for deliberative conversations about the use of AI in science. These forums and engagements must combine periodic convenings with shared channels of communication, including listservs and open digital workspaces. Importantly, these forums should not function as one-off events, but as ongoing spaces for collective inquiry and shared experimentation: places to surface and frame urgent questions about how AI should and should not be used in science, to identify points of disagreement, to test emerging norms and standards, and to explore how those standards can be incorporated into research practice, funding, and policy.

AI is a fast-moving technology, and there is much we do not know. In parallel with forums for discussion, growing the evidence base is critical. Greater investment in and coordination of metascience and AI evaluation focused on scientific and research tasks can help build this base. But funding research isn't enough. To be most

AI x Biodesign, which are already having dedicated conversations around AI in research, can coordinate efforts with like-minded communities from other fields. Organizers of metascience conferences such as the International Conference on the Science of Science and Innovation and AI research conferences like the AAAI Conference on Artificial Intelligence and Conference on Neural Information Processing Systems can carve out facilitated sessions for practical reflection and discussion. Federal funders can sponsor projects and create programs to better understand researchers' AI use and attitudes toward AI, as well as help build the evidence base. Our own convening was supported by the National Science Foundation's Office of Advanced Cyberinfrastructure.

Such efforts need not be overseen by any single entity but must be coordinated and scientist-led. The scientific community has come together in this way before, from historical cases such as the 1975 Asilomar Conference that informed the trajectory of biotechnology, to interdisciplinary reckoning with reproducibility and replicability in recent years, to NASEM efforts

## Stakeholders across the research and development ecosystem need sustained, shared forums for deliberative conversations about the use of AI in science.

effective, these activities must be integrated with forums and discussions with structured opportunities to share evidence, compare approaches, and accumulate learning over time. Together, such efforts can help build a more coherent, evidence-informed approach to AI in science that reflects shared responsibility, distributed expertise, and a commitment to rigor, transparency, and validity.

The health of the scientific enterprise requires “speed bumps”—time and space to ask questions, reflect, discuss, and debate—to challenge the mindset that rushes to adopt AI for science. This responsibility for a more deliberate approach is distributed across the community. At the most integrative level, cross-disciplinary scientific societies and leadership organizations such as the National Academies of Sciences, Engineering, and Medicine (NASEM), American Association for the Advancement of Science, and American Academy of Arts and Sciences have experience bringing together diverse stakeholders and are well positioned to drive sustained convening efforts and nurture emerging consensus. Nonprofits like ORCA can convene creative coalitions to work on specific problems. Communities like Responsible

that influenced the design and passage of the America COMPETES Act in 2022. In this new moment for science, the community must provide coordinated venues to discuss and shape how AI-enabled science is done.

Right now, important conversations are happening in company conference rooms, during conference breaks, on ad hoc panels, or in asides as research projects are conducted. But grappling with the many ways that AI is changing science at this moment must not be confined to the interstitial spaces. The choice is no longer whether AI will reshape both science and its relationship with society. Rather, it is whether the scientific community will play an active and strategic role in how this phenomenon plays out, or whether we will simply watch it unfold from the margins.

*Catherine Aiken is the director of data science and research at Georgetown University's Center for Security and Emerging Technology (CSET). Steph Batalis is a research fellow studying the intersection of AI and life science at CSET. Greg Tananbaum is the executive director of the Open Research Community Accelerator.*