Thinking & Doing Science in the Age of AI

Alison Noble

We are in a new era of advancing science through AI. But the tremendous progress is not just about technology: as scientists have adopted AI-based technologies, they have changed how they do science, redefining the role of human intelligence and the human scientist in the scientific endeavor. This essay reflects on some of these changes and some of the current challenges and tensions that science and scientific communities are facing as we grapple with how best to work with AI to advance science and ensure society's continued trust in scientists and AI-based scientific evidence.

As an engineering scientist who trained in computer vision and has subsequently worked in medical image analysis for nearly thirty years, it has been a privilege to observe and experience firsthand how, in a remarkably short timeframe, artificial intelligence–based techniques have advanced an applied-science discipline. It is hard to believe that it was just over a decade ago (or three generations of UK PhD students) that the first computer-vision methods achieved (average) human-level performance at automated object recognition in one of the early computer-vision data challenges.[1] No longer just a curiosity of computer-vision researchers working at the interface with medical imaging and radiology, there are numerous examples today of automated and assistive AI-based methods for disease detection, anatomical segmentation, and disease characterization that have been translated from concept to clinical research tools and regulatory-approved medical diagnostics used in health care settings. Clinical AI has emerged as a very active area of academic clinical medicine, with its own dedicated journals and conferences; researchers study issues around accuracy, bias, safety, and explainability of deployed AI-based methods and their readiness to be adopted safely into workflows.

Transformations of this kind are mirrored in a number of data-driven science disciplines in which the pattern recognition capabilities of AI can match, or in some cases exceed, human-level capabilities. The 2024 Nobel Prize in Physics, awarded in recognition of artificial neural network methods, and in Chemistry, for using AI systems to create novel proteins and to solve the protein-folding prediction problem, represented a "coming of age" of AI tools for scientific discovery.[2] Significant AI-related advances are also being made in other data-rich areas of science, such

https://doi.org/10.1162/DAED.a.989

as large-scale climate and public-health predictive modeling, and in other areas where research communities are pooling together and releasing large-scale datasets, such as astrophysics.[3] Elsewhere, AI is being introduced as a productivity tool to accelerate real-world workflows and streamline experimental laboratory processes and analyses. Examples range from robotic chemistry lab assistants that can, in theory, work around the clock, video analytics for ethology and conservation, and AI software coding tools for software development.[4] In these cases, the speed at which the AI assistant works has been key to their adoption and success. Much of science prior to the current AI era has traditionally been associated with reflection, not speed. So is AI-based science now challenging the traditional model of scientific advancement, or is it just an ever-expanding suite of new tools that automate existing procedures and accelerate discoveries in existing experimental data? Should we worry that artificial intelligence will replace human intelligence and critical thinking in science, or embrace change and adapt? The answer to the latter question is probably a bit of both.

Because many AI systems require enormous volumes of training data, the earliest adopters of AI in science have, perhaps predictably, come from data-driven science fields that have ready sources of high-quality and well-curated large datasets and that have employed highly skilled computational scientists working alongside domain experts in a small team. Effective as this has been, it is not a scalable approach across the whole of science; there simply are not enough highly skilled computational scientists to meet demand. Training domain experts in AI is also not trivial, and approaches to teaching non-AI scientists essential AI skills are continuously evolving with the AI techniques themselves. Today, for instance, I would argue that a "science and AI" (SciAI) scientist does not necessarily need to be able to program a deep-learning architecture from scratch because of the increasing availability of open access AI models. But they do need to understand the principles of design, build, and testing for the AI-modeling techniques they are using, as well as how to place use in the context of the science. They also need to understand how to rigorously evaluate AI models and challenge their potentially erroneous outputs. This is where human-in-the-loop science domain knowledge and skill are still paramount.

Data, and in particular access to high-quality and well-curated large datasets, are essential to ensuring that AI models accurately and reliably represent the scientific problems they are designed to address.[5] The complexities of data management, curation, and stewardship should not be underestimated; for many researchers, obtaining approval to share and access data – essential for building models and reproducibility – is currently one of the greatest barriers in SciAI research. Experts on data management and data governance are in high demand and become valued members of science and AI research teams. Good data governance is required to ensure that data used to train and test AI models are used responsibly and ethically.

Best practices in scientific data governance and management of data privacy and copyright lag behind AI algorithms and are complicated by regional, national, and sector-specific legal requirements. Data governance for medical research use may, for example, have to satisfy GDPR (General Data Protection Regulation) requirements in Europe as well as HIPAA (Health Insurance Portability and Accountability Act) rules in the United States. The legal, regulatory, and ethical issues associated with data governance are particularly challenging for global science collaborations, in which researchers need to navigate different national and international data regulations that are evolving at different rates.

Some science communities have for decades been working collaboratively on data curation and on construction of open-data challenges. The European Molecular Biology Laboratory European Bioinformatics Institute (EMBL-EBI) is a good example. In 1994, the EMBL-EBI set up the Critical Assessment of Protein Structure Prediction (CASP), an international competition for computational model prediction of protein folding, which was won by the AlphaFold2 AI system at CASP14. Meanwhile, organizations such as Hugging Face and Kaggle have provided platforms to enable research communities to collaborate and benchmark computational methods on open datasets. With sensitive data areas like health care, research is shifting from a data-sharing model, whereby raw data are transferred between partners typically covered by a data-sharing agreement, to a data-access model, which has tighter controls on data privacy and security, with data remaining at the source of data generation.

One data-access approach allows AI analysis to take place only in a secure data-analytics environment. Data-access and compute charges are needed to cover the running costs and sustainability of such research infrastructure, and debates on fairness and equity have been ongoing for some time, as fee models favor industry and other well-funded groups and are out of reach for many early-career researchers and researchers with limited resources.

For groups that wish to collaborate to train a single model but cannot share data between sites, federated learning may provide a solution. In this case, a (global) model is sent to participating clients (for instance, where each has a radiology imaging database), each client uses its local, sensitive data to train and improve the model locally, and the client sends updated model parameters, but not the data, back to the centralized system, which then aggregates updated model parameters for all clients to give the updated global model. This approach has been applied in COVID-19 analysis, for instance, but there have been few other real-world scenario examples of research use so far.[6] This is because of outstanding theoretical challenges, including how to effectively model tasks for the "small community experiment" of a small number of clients and highly heterogenous, possibly multimodal data; and practicalities, including computational resource requirements at client sites. Open-source tools, such as Flower, that provide the infrastructure

for implementing federated learning–based modeling are gaining traction at the current time.[7]

Finally, some research areas (such as computer vision and large language models) have traditionally used large training datasets generated by (sometimes indiscriminately) scraping the internet for data. This sourcing raises questions about data copyright. But there is currently no international consensus on how to deal with copyrighted data used to train AI models for scientific research, highlighting the complexities and tensions between the intellectual property rights of the data author, open science, and the advancement of science by AI.

It is important to embrace the positives that the new AI tools and capabilities offer for scientific breakthrough, innovation, and potential economic and social impact. But we must also recognize that the use of AI tools in science challenges some of the established norms of scientific research – and that research ethics has never been more important in science than today.

Responsible AI (RAI) is central to trustworthy, safe, and ethical development and use of AI-based scientific methods. Community-led initiatives, including a collaboration between the National Academy of Sciences, the Annenberg Public Policy Center of the University of Pennsylvania, and the Annenberg Foundation Trust at Sunnylands, and work by UNESCO and Responsible AI UK have set out to identify AI-expert consensus in this area and to encourage adoption of RAI principles and best practices.[8]

Scientific rigor and transparency matter. They are foundational principles of science as an evidence-based discipline. All AI models will make errors, typically due to limitations of the training data (such as data distribution bias, data size, or gaps in representational completeness of a problem) and the model itself (observed as parameter over- or underfitting). Why should we worry about this? After all, these are not new issues in scientific data modeling, and there are ways to identify and in some cases mitigate these factors. However, the opaqueness (black-box nature) of AI models makes it easier for researchers to accept an AI-generated conclusion when it is wrong, particularly if it gives a result that matches their hypothesis (the so-called echo-chamber effect in critical thinking). The current trend of preprint publication of scientific research without peer-review of the article is not helping either. The extent to which this is leading to false claims and publication of work that cannot be reproduced is worrying scientific communities. In the context of AI, good scientific practice requires researchers to rigorously design, implement, and test a model to ensure reproducibility, and to rigorously analyze bias, accuracy, safety, and explainability (BASE). Peer reviewers of articles featuring AI models cannot verify the models by reproducing them, so they depend instead on trusting the authors' descriptions of methods and (where appropriate) benchmarking with other published models on common data. Yet surprisingly few science

journals or conferences currently insist on rigorous evaluation and reporting, typically just requiring reporting on AI model accuracy and (sometimes) explainability. Beyond the components of BASE, ideally all scientific AI work should be reported according to FAIR (findable, accessible, interoperable, reusable) principles, as well as cover descriptions of datasets and encourage code and data to be made open for other researchers to use for benchmarking, where possible.

To address these concerns, comprehensive consensus-based guidelines and standards of reporting have been proposed in some disciplines, including CLAIM (AI in medical imaging) and REFORMS (general machine learning–based science).[9] Adoption of such checklists is relatively new; it is important that these initiatives are given high visibility and community-wide support to improve standards of reproducibility and transparency, including by embedding them in the training of SciAI scientists.

The degree to which open science principles have been adopted in different scientific disciplines to date varies significantly. Communities that have traditionally worked together on data curation and data and code sharing have found this transition easier; many work together on data challenges. Those who work with sensitive real-world data that typically have restricted access, as with AI in medical imaging, follow a mixed model of working with public and private datasets. This creates a tension between researchers who insist that new method assessment should include analysis of public data and researchers who model novel tasks with novel private data, for which a public data equivalent does not exist. Following FAIR principles is hardest for disciplines that have traditionally comprised siloed groups, whose datasets are their assets and competitive edge. Time will tell if the challenges of reproducibility in AI and the desire to use AI to answer new scientific questions in such areas will shift attitudes and behavior from competition to partnership.

As with reproducibility, recent advances in AI, and particularly when generative AI (GenAI) is used as part of the research method, are requiring new thinking about the ways we look at scientific integrity. GenAI can generate synthetic scientific data, including images and text, and perform tasks such as text summarization, automated code generation, and image editing. As AI-generated (synthetic) content becomes more realistic and more prevalent, it must not be confused with real-world observations. In science, GenAI in its various guises, including large language models (LLMs), offers many advantages, ranging from literature synthesis to experimental design to writing, but it needs to be used with care. For this reason, an interdisciplinary panel convened by the National Academy of Sciences proposed five principles of human accountability and responsibility when using GenAI in science: 1) transparent disclosure and attribution, 2) verification of AI-generated content and analyses, 3) documentation of AI-generated data, 4) ethics and equity, and 5) continuous monitoring, oversight, and public engagement.[10] The need for ethical

standards in AI-assisted writing and greater transparency in LLM use is also discussed, for instance, in recent work by bioethics scholar Sebastian Porsdam Mann and colleagues.[11] Issues of attribution also extend to the patenting of inventions involving AI. Patents are typically awarded for human creativity and currently there is no clear guidance on how to manage co-inventions between humans and AI.

No discussion of AI in science can ignore how AI is changing the future of scientific publishing. Throughout the history of science, scientists have communicated to their peers and the public in continuously changing ways, from peer-to-peer letter writing and public lectures in the early days to today's potpourri of dissemination methods for what are often global audiences, including peer-reviewed articles, databases and code repositories, conferences, and podcasts and web videos, to name a few. The scientific publishing industry is itself undergoing significant changes to incorporate AI and automate processes, but the discussion here will focus on perspectives related to the scientist (as the customer) and on how LLMs – the special case of generative AI that focuses on language-related tasks and is adept at search and mimicking human writing – are impacting science article content and article peer review.

The emergence of LLMs that can perform routine search and summarization/content discoverability at remarkable speed and scale means the end of literature and systematic reviews as we know them today. Many researchers take this as good news because it frees up their time for more creative tasks. Others argue that while current LLM-based summarization tools can capture an article's story, findings, and impact, they cannot assess the trustworthiness of code, methods, results, and analysis as well as people. If so, then humans still have roles in checking the AI assistant's results and performing original synthesis – though that may not be for long, depending on the literature reviewing task and how the technology develops. This is also an area where emerging human-AI collaboration technologies such as learning-to-defer, in which AI performs a task but defers to a human when the AI is unsure, may play a useful role.[12]

For peer review, the challenge and opportunity are to figure out how best to share the burden between AI and humans while ensuring the quality of the review process. There is already a wide spectrum of ways to publish scientific articles, from the preprint model of publishing first and discussing later to traditional peer-review of articles submitted to journals and conferences. These approaches serve different purposes in science (to share scientific findings with others, to create a scientific record, to gain professional recognition and influence) and we should respect and celebrate this diversity. However, it is well acknowledged that in many science disciplines, the current journal peer-review system is broken: the volume of papers submitted and demand for review exceeds human scholar availability. Additionally, human peer review can be prone to bias and can struggle to detect fraudulent work. Hence there is significant interest from publishers and scientists in understanding

what role LLMs can play (or not) in supporting the review process and which types of scientific work LLMs can assess. Scientists have different views on this. Moreover, as science journalist Miryam Naddaf has noted, it is critical to set clear standards of transparency so both the authors submitting articles for review and readers of the journal are informed that AI is part of the review process.[13]

If, in this new era of science and AI, our notions of scientists and the scientific method are changing, then what is the role of human intelligence and the human scientist?

While there are recent papers presenting early ideas and demonstrations of AI systems (agentic and nonagentic) that can automate entire scientific discovery processes, the AI tools we will increasingly see in AI-based science for the foreseeable future are assistive-AI systems (narrow AI) tuned to the needs, and built on the data, of a specific science domain.[14]

If AI systems can meet the (high) expectations of scientists and can be accepted as a team member in a science laboratory, does this change our expectations of the role and skill set of human scientists? These changes relate to how humans and AI can effectively work together, and who does what in the thinking and doing of science.

Where AI fully automates processes, such as in an experimental laboratory, jobs for humans will undoubtedly disappear. Elsewhere, the breadth of skills required by a human SciAI scientist are increasingly becoming an interdisciplinary mix of science domain expertise and knowledge, AI, and research ethics. For instance, researchers using AI tools in classic experimental disciplines such as biology, physics, and chemistry may now need to master data (analytical) skills, blurring the distinction between a computational scientist and experimental scientist of days gone by. Many AI tools are AI assistants, so the human scientist's role includes challenging the AI to verify outputs.

A decade ago in my own research area, medical image analysis, we would have labeled medical imaging AI technologists working with clinicians as interdisciplinary. Today, we also work with experts in human factors and with experimental psychologists to understand not only the accuracy and reliability of the systems we develop but to study, for instance, how human trust is essential for end user acceptance and safe deployment of those systems.[15] This means the subject is becoming truly interdisciplinary, combining computer science, engineering, medicine, and the social sciences.

Do individuals necessarily need an interdisciplinary mix of skills? Probably not. And there are lessons academia might learn from the private sector in terms of how to effectively conduct SciAI research and development, by which a team of computer scientists, high-performance computing engineers, data managers and data governance specialists, science domain experts, and social scientists co-design, -build, and -evaluate models. Building scientific teams of this disciplinary

breadth and depth comes at a cost, but where the potential scientific gain is great, this model of interdisciplinary scientific collaboration, including public-private sector partnerships, is likely to become more common.

As a final point, I want to comment on recent research on the effect of AI on critical thinking. Critical thinking is a fundamental cognitive skill and an essential part of problem-solving and reflective thinking in scientific work. A recent study by AI and society scholar Michael Gerlich found a significant negative correlation between frequent AI tool use and critical thinking.[16] This suggests that as we move forward with more AI integration into the scientific method, we need to ensure that the benefits of cognitive offloading and the desire to be first in AI-based scientific discovery are not at the expense of critical thinking. If we do not, human scientists risk losing the habit of thinking. While AI tools may assist with the doing, the thinking in science and AI needs to remain the responsibility of the human.

Artificial intelligence technologies are supporting scientists in many disciplines to advance scientific discovery and improve productivity. Our responsibility as scientists is to ensure rigor in our standards of reproducibility of scientific evidence, the scientific integrity of AI-based work, and that human critical thinking stays at the heart of science advancement. The rapid adoption of new AI technologies for task automation and human-AI collaboration to empower human scientists is all but inevitable. While scientists are rightly concerned about how AI technologies work and behave, understanding the human element of human-AI collaboration in SciAI research may be the greater challenge.

ABOUT THE AUTHOR

Alison Noble is the Technikos Professor of Biomedical Engineering at the University of Oxford. She is also Vice President and Foreign Secretary of the Royal Society and Chair of the Royal Society working group that produced the policy report *Science in the Age of AI* (2024). She previously served as Director of the Institute of Biomedical Engineering and as an Associate Head of Division at the University of Oxford. She has recently published in such journals as *Nature Biomedical Engineering*, *npj Digital Medicine*, *Medical Image Analysis*, and *Proceedings of the AAAI Conference on Artificial Intelligence*.

ENDNOTES

[1] Olga Russakovsky, Jia Deng, Hao Su, et al., "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision* 115 (3) (2015): 211–252.

[2] John Jumper, Richard Evans, Alexander Pritzel, et al., "Highly Accurate Protein Structure Prediction with AlphaFold," *Nature* 596 (7873) (2021): 583–589, https://doi.org/10.1038/s41586-021-03819-2.

[3] Multimodal Universe Collaboration (Eirini Angeloudi, Jeroen Audenaert, Micah Bowles, et al.), "The Multimodal Universe: Enabling Large-Scale Machine Learning with 100TB of Astronomical Scientific Data," arXiv (2024), https://doi.org/10.48550/arXiv.2412.02527.

[4] Tianwei Dai, Sriram Vijayakrishnan, Filip T. Szczypiński, et al., "Autonomous Mobile Robots for Exploratory Synthetic Chemistry," *Nature* 635 (8040) (2024): 890–897, https://doi.org/10.1038/s41586-024-08173-7; and Max Bain, Arsha Nagrani, Daniel Schofield, et al., "Automated Audiovisual Behavior Recognition in Wild Primates," *Science Advances* 7 (46) (2021), https://doi.org/10.1126/sciadv.abi4883.

[5] National Academy of Sciences, *Toward a New Era of Data Sharing: Summary of the U.S.-UK Scientific Forum on Researcher Access to Data* (National Academies Press, 2024), https://doi.org/10.17226/27520.

[6] Ittai Dayan, Holger R. Roth, Aoxiao Zhong, et al., "Federated Learning for Predicting Clinical Outcomes in Patients with COVID-19," *Nature Medicine* 27 (10) (2021), https://doi.org/10.1038/s41591-021-01506-3.

[7] See Flower AI, https://flower.ai.

[8] Wolfgang Blau, Vinton G. Cerf, Juan Enriquez, et al., "Protecting Scientific Integrity in an Age of Generative AI," *Proceedings of the National Academy of Sciences* 121 (22) (2024), https://doi.org/10.1073/pnas.2407886121.

[9] Ali S. Tejani, Michail E. Klontzas, Anthony A. Gatti, et al., "Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 Update," *Radiology Artificial Intelligence* 6 (4) (2024), https://doi.org/10.1148/ryai.240300; and Sayash Kapoor, Emily M. Cantrell, Kenny Peng, et al., "REFORMS: Consensus-Based Recommendations for Machine-Learning-Based Science," *Science Advances* 10 (18) (2024), https://doi.org/10.1126/sciadv.adk3452.

[10] Blau, Cerf, Enriquez, et al., "Protecting Scientific Integrity in an Age of Generative AI."

[11] Sebastian Porsdam Mann, Anuraag A. Vazirani, Mateo Aboy, et al., "Guidelines for Ethical Use and Acknowledgement of Large Language Models in Academic Writing," *Nature Machine Intelligence* 6 (2024): 1272–1274, https://doi.org/10.1038/s42256-024-00922-7.

[12] Hussein Mozannar and David Sontag, "Consistent Estimators for Learning to Defer to an Expert," arXiv (rev. 2021), https://arxiv.org/abs/2006.01862.

[13] Miryam Naddaf, "AI Is Transforming Peer Review–and Many Scientists Are Worried," *Nature*, March 26, 2025, https://www.nature.com/articles/d41586-025-00894-7.

[14] Chris Lu, Cong Lu, Robert Tjarko Lange, et al., "The AI Scientist: Towards Fully Automated Open-Ended Scientific Discovery," arXiv (2024), https://doi.org/10.48550/arxiv.2408.06292; and 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.

[15] Angus Nicolson, Elizabeth Bradburn, Yarin Gal, et al., "The Human Factor in Explainable Artificial Intelligence: Clinician Variability in Trust, Reliance, and Performance," *npj Digital Medicine* 8 (2025): 658, https://doi.org/10.1038/s41746-025-02023-0.

[16] Michael Gerlich, "AI Tools in Society: Cognitive Offloading and the Future of Critical Thinking," *Societies* 15 (1) (2025): 6, https://doi.org/10.3390/soc15010006. See also the references therein.