

Commentary

Artificial Intelligence for **Scientific** Discovery: Is Europe Ready for It?



[David Osimo](#)



[Jonathan Wareham](#)

The disruptive opportunities of artificial intelligence (AI) in science are already visible, both in academia and industry. AI enables simulations at an unprecedented scale, using radically bigger amounts of data. It allows for automated experimentation and faster time to market new products and services. It is accelerating discovery at an unprecedented rate, as we saw with the COVID-19 vaccine, reducing the time needed for clinical trials. Analysts at McKinsey Global Institute predict research and development as one of the main areas of impact, with productivity gains between [10% and 15% from AI adoption](#). As such, AI is expected to contribute to making research more productive and effective. Therefore, as the European Commission illustrated in [Harnessing the Power of AI to Accelerate Discovery and Foster Innovation](#), a recent policy brief, it is paramount to get the policy framework right if Europe wants to remain at the forefront of science and innovation.

One of the most successful beacons of AI in science is [AlphaFold](#). [Google DeepMind](#) and the [European Molecular Biology Laboratory of the European Bioinformatics Institute \(EMBL-EBI\)](#) developed AlphaFold, a neural network system, to predict the structure of proteins – a long-standing challenge in biology and pharmacology. When done through experimental methods such as X-ray crystallography, mapping a single protein can take several years of dedicated research. Thanks to this breakthrough collaboration, AlphaFold consistently matches the prediction accuracies of experimental methods to determine the three-dimensional structure of proteins at far greater speeds. In 2022, AlphaFold published predictions for almost all known proteins: 200 million proteins compared to the 200,000 that were thus far mapped using experimental methods. The results were published in an open database hosted by EMBL-EBI, which has already been accessed by 1.4 million researchers and used in very diverse applications such as [antibiotic resistance](#) and [the fight against plastic pollution](#). It is expected that AlphaFold will fuel a radical acceleration in the time it takes to develop new drugs, saving millions of research years – and possibly even millions of lives.

The case shows the importance of open and collaborative science. AlphaFold was trained on open public data, and its predictions released in an open-access database. Public-private collaboration

was necessary not to generate the data but to curate and fully understand how to design a database that is useful for science and industry. Computational biologists and AI experts from DeepMind and EMBL-EBI collaborated closely and iteratively.

Interestingly, AlphaFold also embodies a fluid relation between theory and experiment in scientific AI. Throughout its computations, AlphaFold is unconstrained by extant theory until the final phase where the results are assessed against known laws of chemistry and physics. Counterintuitively, including theoretical constraints earlier into AlphaFold's calculations [made the algorithm less effective](#). The fact that AlphaFold can be more successful when allowed to violate chemical laws is discomfiting – a loose relation with theory has traditionally been hard to accept in scientific research – and remains polemical to this day. The best-known early example of atheoretical scientific computation was the Monte Carlo simulations used in post-World-War-II research on the hydrogen bomb. As a pseudo-random number generator, it was developed to model the stochastic processes of nuclear fusion that were inaccessible both experimentally and mathematically. Because of its black box inscrutability, such atheoretical computations have long been viewed with scepticism. Yet, over time, theory-free numerical simulations and machine learning have become essential tools across many fields of science: A tension born in post-WW-II physics that is taking many new turns and today's AI science, as we show in a recent [paper](#).

Yet, AlphaFold is just one success story.

Does Europe, as a whole, have what it needs to excel in AI for science?

The basic tools are already available to all researchers to use, but new models are expected to emerge continuously and require continuous adaptation. Europe is well-positioned to lead. It is well-equipped in terms of infrastructure, as it hosts three of the top 10 supercomputers in the world. Its human capital is among the most advanced globally, with more than two million researchers in 2022, up 45% from 2012. Yet, the global competition is fearless: China has overcome Europe in the volume of scientific publications, while the United States remains the leader in terms of quality and impact. China and the United States outperform Europe in terms of fast-growing companies, also known as unicorns.

This competition is only stronger when it comes to AI adoption in science. As the European Commission points out in its *Harnessing the Power* report, Europe is becoming less attractive for AI researchers, as "20% of top European Union AI researchers went to the United States for their graduate school and a further 14% left for their post-graduate work." In addition, current data paint a worrying picture in terms of the adoption of AI by scientists in Europe compared to other regions. China has overtaken the European Union and the United States in terms of both the number and share of publications on AI applications in science.

To be clear, the opportunities to adopt AI in science will not come from a large language model chatbot substituting scientists. The next discoveries of humankind will not be generated by prompt engineering but by close collaboration between computational scientists and field-specific expertise. And this collaboration is complex, requiring cross-disciplinary understanding, immense computing resources, iteration and trust between different players. AI is not a ready-made solution that can be lapped on as a plugin. It requires human oversight, extensive tweaking and trial and error to fine-tune the models and ensure their usefulness.

In summary, for Europe to retain its leading role in research and innovation, it needs to ensure it is at the forefront of developing and using of AI. And this requires an ambitious strategic approach. Getting the policy framework right is therefore paramount and many countries are working on it by developing both targeted and framework policies.

Targeted research and innovation (R&I) policies typically focus on data, skills and infrastructure. AI needs high-quality, well-curated data, such as those that EMBL-EBI holds. It needs skilled AI developers and computational scientists across different disciplines. And it needs state-of-the-art supercomputing infrastructure to train the models. In this sense, the proposals in the recent European Commission policy brief are in the right direction.

But targeted R&I measures are only part of the story. Framework policies can be even more important. In this sense, there is a concern that the ambitious regulatory effort the European Union has undertaken might harm the large-scale adoption of AI, in particular, because of the strong focus on *ex ante* measures as guardrails for the unpredictable and potentially harmful outcomes of AI. As the authors argue in "[Regulating AI: Lessons from Scientific Computing](#)," a recent article co-written by four scientists, the AI Act focuses much more on detecting the risks before an AI deployment is launched into the market rather than post-market surveillance. Appropriately, it stresses provisions for adequate risk assessment and mitigation, high-quality datasets, traceability, documentation, transparency, human oversight, security and accuracy among others. These are all desirable objectives, particularly in the high-risk applications of AI. Yet, at the same time, AI's inherent inscrutability is not exclusively a risk or liability to be confined and feared. Much of AI's potential value lies in letting it address problems with its own logic, unconstrained by extant rules or any specific theoretical apparatus. This is clearly the case in many scientific applications such as AlphaFold. As such, effective AI regulations should not only be focused only on its potential risks, but also on nurturing its benefits – allowing it to shine where it is strong – with a pragmatic balance of *ex ante* and *ex post* measures.

Of course, the AI Act includes an exception for models exclusively developed for research and innovation purposes. But it remains to be seen whether these kinds of procedural exceptions work in a digital market where the boundaries between science and various industrial sectors are increasingly blurred. Many of the recent trepidations about AI transparency reflect past debates on the adoption of computational methods in science; concerns that were often eclipsed by the usefulness of the results. A lot will depend on the implementation of the regulation and on the culture that surrounds it. A simple belief in the need to be in full control of the technology and fully understanding how it works *ex ante* can hinder the realisation of AI's tremendous potential and prevent Europe from maintaining its leadership in science and innovation.

David Osimo is director of research at the Lisbon Council, a Brussels-based think tank. Jonathan Wareham is professor of information systems at ESADE business school.

theForum

The Forum is a semi-weekly newsletter of cutting-edge blogs devoted to policy proposals and up-to-date political analysis. It is hosted by the Lisbon Council, a global think tank based in Brussels. For more commentary visit: <https://lisboncouncil.net/category/blog/>