The Influence of Artificial Intelligence on Scientific Knowledge and the Role of the Philosophy of Technology: A Philosophy of Science Perspective[†]

Florian J. Boge, Institute for Philosophy and Political Science, TU Dortmund University & Lamarr Institute for Machine Learning and Artificial Intelligence

1. Introduction

Artificial intelligence (AI) is omnipresent today. Science, too, is increasingly permeated by AI: in 2024, not just one but two Nobel Prizes went to researchers who had made significant contributions to, or with the help of, AI. But what makes AI so special, especially in science? Important aspects are certainly the speed and scope of the developments it drives forward. Thus, for example, the 2024 Nobel Prize in Chemistry was awarded to John Jumper and Demis Hassabis for developing the protein structure prediction model *AlphaFold2*, which partly uses the same basic principles as the language model ChatGPT.

Proteins are ultimately curled-up and specifically folded chains of amino acids. There is usually a close and stable relationship between an amino acid chain and its fold, and the fold often determines the protein's function. With AlphaFold2, it became possible for the first time in 2020 to accurately predict structures based on amino acid chains. This has drastically increased the number of known (or: predicted) protein structures: Alongside roughly 300,000 empirically known structures, there are now over 200 million predicted by AI (Callaway, 2022). Thus, AI has solved a prediction problem that had persisted for fifty years. Nevertheless, questions remain: Why does a protein fold as it does? What are the underlying mechanisms? How could we design proteins ourselves to trigger physiological reactions?

2. AI from a Philosophy of Science Perspective

One pressing question thus is whether AI can promote scientific understanding in addition to scientific prediction. As the case of AlphaFold2 shows, this is far from clear. A plausible thought is that this difficulty stems from the *epistemic opacity* of AI systems such as Neural Networks (NNs) like ChatGPT or AlphaFold2: mathematically speaking, such NNs are complex nestings of linear and nonlinear functions with sometimes hundreds of billions of free parameters.

Emily Sullivan (2022) has argued, in an influential article, that it is not merely opacity itself that hinders understanding. Rather, on Sullivan's view, uncertainty about how these models connect to scientific knowledge—particularly: empirical evidence—is responsible for AI's providing predictions without simultaneously yielding understanding of the world. In Boge (2022), however, I argued that this does not go far enough.

[†] This is a translation of the preprint of an article forthcoming in Friedrich, A., Gehring, P., Nordmann, A., Hubig, C. & Kaminski, A. (eds), *Jahrbuch Technikphilosophie 2025*, Nomos. The final version may differ from this one in several ways.

Specifically, there seems to be a new dimension of opacity unique to AI systems. For instance, in highly complex computer simulations, it may be unclear how they function in detail, what happens during individual runs, or how their output comes about (Creel, 2020; Beisbart, 2021). All this certainly applies to AI as well. But moreover, it remains unclear what exactly an AI "finds out;" i.e., what information it extracts from data that humans can hardly recognize.

For instance, physicists Chang et al. (2018) showed that an NN could infer complex quantities from data, such as masses of intermediate products in decay chains, to classify particle collision data according to whether they corresponded to new or known particles. That is precisely the information which physicists themselves infer to discriminate between different kinds of data, on the basis of physical laws. If one statistically "erases" this information from the data, then the AI fails, indicating that it must have independently discovered those relevant quantities itself. We must therefore recognize another dimension of opacity; also unknown is the answer to the question: What does AI actually learn?

3. What Can Philosophy of Technology Contribute?

Reading all this might suggest that philosophy of technology is irrelevant for assessing AI's significance in science: AI's unique features must be captured purely within philosophy of science, with the aid of concepts from the philosophy of mind (see also Boge, 2024). However, this impression is not entirely correct.

In an article from 2019, Paul Grünke and I examined the "depth" of the opacity associated with AI (Boge & Grünke, forthcoming). Following and extending Paul Humphreys (2009), we proposed a threefold distinction between levels of opacity: (i) A process is *opaque* if not all epistemically relevant elements are known; (ii) *essentially opaque* if one *cannot* know them all; and (iii) *fundamentally opaque* if *no one* can know them all—the third level being our own extension. Another novel idea was to distinguishing sources of opacity: Does it arise due to complexity or is it inherent to method?

Now, does AI in fact add to depth of opacity? The answer: no, all levels can occur independently of AI. Yet, doesn't AI gain its particular opacity precisely because its method involves delegating learning new facts to the machine? The answer here is: Yes, but that methods can be inherently opaque is nothing new either. How can this be shown?

In the paper, we offer a careful analysis based on the use of technology in scientific measurements, as a foundation for our verdict. Why should technology count as epistemically opaque at all? Kaminski (2018, 330), for example, argues that computer simulations only become genuinely epistemically opaque through the entanglement of technology and mathematics—for in principle, every aspect of a device (such as a digital computer) could be fully inspected.

As we show in our 2019 paper, this is only half right. For instance, there are uncertainties in the behavior of measurement instruments quite similar to those regarding the functioning of AI. Thus, even manufacturers typically know only up to a certain degree how an instrument or implement they have manufactured operates in detail. A lot is heuristic knowledge, measurement, or even just practical experience. This effect amplifies when multiple components—often produced by different manufacturers—have to interact. Without such interactions, many instruments cannot function at all; but interaction normally introduces further sources of uncertainty. There does not have to be a very large number of parts for this, though, whence this is not a mere issue of complexity—the relevant sort of opacity is inherent in the method of measurement itself.

Using a detailed case study from high-energy physics, we demonstrate how interpretation depends crucially on such uncertainties, whence researchers must constantly reevaluate, and where possible reduce them. Hence, there are epistemically relevant elements of a measurement which are unknown, and every sufficiently advanced measurement inherits a certain amount of epistemic opacity.

However, with AI, things are thus very similar: Many algorithms are based on heuristic principles, justified through benchmarking successes, or are simply adopted because they work in practice more generally speaking. Why an algorithm works so well, or why it fails in certain instances, is thereby often left opaque.

Moreover, it is often unknown what the effects of interactions among numerous internal components are, each associated with its own uncertainties. This is especially salient in Deep Neural Networks, wherein repeating structures, which might be understandable individually, become nested in such ways that their global interplay becomes opaque as well. This too is not purely an effect of complexity: Even in rather small NNs that could be inspected in principle, it is not initially clear what individual units are doing in the overall functioning of the NN, or how the NN as a whole comes up with its prediction.

Similar difficulties as in the case of measuring devices also arise with respect to "manufacturers": Many mathematical functions used within standard machine-learning libraries come pre-programmed, but are computed only approximately. The exact approximation method may vary, but this normally remains hidden at least from users. Additionally, the specific realization on a physical computer does not exactly match the theoretical mathematical approximation, as translation into computer code, as well as realization on hardware, introduces further changes (see also Boge, 2020). Thus, specific details of "calling up" pre-programmed functions also remain obscure even for their programmers.

As we can see, numerous analogies exist between measurement instruments and Machine Learning, and the latter's opacity stands in some continuity to the formers'. From a philosophy of science perspective, however, a decisive role for the philosophy of technology thus emerges: It may help us to understand AI as A new technology, and to sort out and grasp its special features in comparison with other technologies. Once we possess general insight into how technological epistemic opacity arises and what it implies for scientific research, we can certainly also transfer such insights into philosophical analysis concerning use of AI within science and beyond.

References

Boge, F. J. & Grünke, P. (forthcoming). Computer Simulations, Machine Learning and the Laplacean Demon: Opacity in the Case of High Energy Physics, in Resch, Kaminski, and Gehring (Eds.), *The Science and Art of Simulation II,* Springer. Online: https://philsci-archive.pitt.edu/17637/

Boge, F. J. (2020). How to infer explanations from computer simulations. *Studies in History and Philosophy of Science Part A*, 82: 25-33.

Boge, F. (2022). Two dimensions of opacity and the deep learning predicament. *Minds and Machines*, *32*(1), 43–75.

Boge, F.J. (2024). Functional Concept Proxies and the Actually Smart Hans Problem: What's Special About Deep Neural Networks in Science. *Synthese*, 203, 16.

Beisbart, Claus (2021). Opacity thought through: on the intransparency of computer simulations. *Synthese*, 199 (3-4):11643-11666.

Callaway, E. (2022). What's next for AlphaFold and the AI protein-folding revolution. *Nature*, *604*, 234–238.

Chang, S., Cohen, T., & Ostdiek, B. (2018). What is the machine learning? *Physical Review D*, 97(5), 6.

Creel, Kathleen A. (2020). Transparency in Complex Computational Systems. *Philosophy of Science*, 87 (4):568-589.

Kaminski, A. (2018). Der Erfolg der Modellierung und das Ende der Modelle:Epistemische Opazität in der Computersimulation, in A. Brenneis et al. (Hrsg.), *Technik – Macht – Raum, Technikzukünfte, Wissenschaft und Gesellschaft*, 317–334. Wiesbaden: Springer.

Sullivan, E. (2022). Understanding from Machine Learning Models. *The British Journal for the Philosophy of Science*, 73:1, 109-133.

Humphreys, P. (2009). The philosophical novelty of computer simulation methods. *Synthese*, 169(3):615–626.