swiss physical society

Nr. 77, November 2025

SPG Mitteilungen
Communications de la SSP

Auszug - Extrait

PT 01/2025

Physical understanding in the times of AI: A philosophical analysis

Claus Beisbart, Universität Bern, Institut für Philosophie, Länggassstrasse 49a, CH-3012 Bern, Claus.Beisbart@unibe.ch

This article has been downloaded from:
https://www.sps.ch/articles/plenary_talks/

PT 01/2025

Physical understanding in the times of AI: A philosophical analysis

Claus Beisbart, Universität Bern, Institut für Philosophie, Länggassstrasse 49a, CH-3012 Bern, Claus.Beisbart@unibe.ch

These days, AI applications, particularly artificial neural networks (ANNs), are all the rage in physics and beyond. Undoubtedly, they can be powerful in classification and prediction tasks. For instance, Lam et al. (2023) have recently developed a neural network model called GraphCast, which outperforms traditional simulation-based numerical weather prediction. However, how far can the use of AI be pushed in physics? In particular, can AI yield scientific understanding of real-world phenomena such as turbulence or superconductivity?

Some authors have been skeptical, suggesting that big-data-oriented science remains shallow. For instance, Kaznatcheev (2013) has remarked that "[s]cientists (due mostly to science funding) start to move away from trying to explain and understand phenomena and more to collecting data and data-mining it to make predictions without understanding." A reason for skepticism is that ANNs themselves are opaque and thus difficult for humans to understand (e.g., Burrell 2016). Others, by contrast, have taken a more optimistic stance, pointing to examples in which AI has seemingly been instrumental to human understanding. In what follows, I'll try to investigate more systematically to what extent AI can help human physicists to understand real-world target systems (see Beisbart 2025 for details).

The answer to this question crucially depends on what we mean by "understanding" or "providing understanding." Although physicists often pride themselves on having deepened our understanding of a particular domain, they frequently do not explain what they mean by "understanding." This is where philosophy of science becomes essential. In the last few years, philosophers such as Catherine Z. Elgin (2017) or Henk de Regt (2017) have clarified what understanding might be. Particularly, philosophers have distinguished between several varieties of understanding, e.g., the (so-called objectual) understanding of some domain of things or the (so-called explanatory) understanding of why something is the case (see Baumberger et al. 2016 for an overview). In what follows, I focus on explanatory understanding, which is particularly important for science. It is widely agreed that this understanding is not just a subjective feeling of understanding (Trout 2002). Instead, if a scientist understands why p is the case, where p is a proposition describing a phenomenon, e.g., a thunderstorm, they are committed to an explanation of why p, which is at least roughly correct. Furthermore, they are justified in accepting the explanation and can use it in scientific practice. For instance, they can formulate the explanation and use it to make inferences about scenarios in which some details in the explanation have been changed (see Hills 2016 for details).

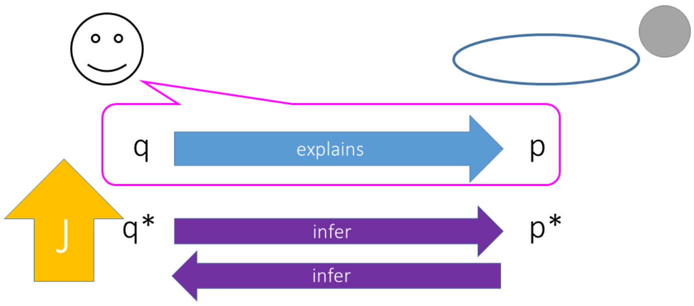

Fig. 1: Sketch of the conditions required for explanatory understanding of why p is the case (e.g., why the planets move on particular orbits). The agent on the left hand side has to be committed to a correct explanation q of p; they need justification (J) for this, and they are required to have the ability to draw inferences about scenarios in which p and q have been modified slightly to p or q*, respectively.*

Suppose now that a scientist has trained an ANN to predict the weather and assume that the ANN's predictions are highly accurate. Does this achievement provide the scientist with an understanding of why the weather is as it is, e.g., why there is a thunderstorm on a particular day? An immediate problem is where the explanation is. Typical physical explanations are based on laws of nature or theories of broad scope; they cite causes of a phenomenon or specify a mechanism to explain the phenomenon under investigation (see e.g. Woodward & Ross 2021). However, it is unclear whether ANNs do this. It is, for sure, possible that they become sensitive to variables that describe the mechanisms underlying the phenomenon. Their inner layers may also contain nodes that trace the values of variables that arise in the relevant laws of nature. However, without further investigation of the ANN, physicists do not know whether this is the case. Accordingly, they don't have an explanation in their hands. As a consequence, they cannot commit to an explanation, and they cannot use it to draw inferences. This means that they lack explanatory understanding.

However, if the ANN yields accurate predictions for a wide range of cases, there are some reasons to expect that the ANN incorporates some explanatorily relevant information.

Still, unless the scientists have identified the explanatory factors that the ANN has detected, they cannot articulate the explanation, let alone use it to run the inferences needed for understanding. This result implies that explainability methods are crucial if scientists want to use ANNs to deepen their understanding: unless scientists understand which explanation an ANN somehow contains, they cannot use it to understand real-world target systems.

Fig. 2. In his "Philosophical Investigations" (Wittgenstein 1953, § 146), philosopher Ludwig Wittgenstein suggested that understanding is not so much a mental process; instead, understanding means to know how to proceed. In this way, Wittgenstein pioneered the idea that understanding manifests in specific abilities. (photograph by M. Nähr, public domain, via https://commons.wikimedia.org/wiki/File:Ludwig_Wittgenstein_1929.jpg)

There is a difference here between ANNs and other computer-based methods, e.g., computer simulations. The latter are typically based on theories that contain the relevant laws of nature, e.g., the basic equations from fluid dynamics, and physicists have invested explanatory knowledge in designing the simulations. ANNs are different because they can be trained without any prior explanatory understanding.

If this is true, why are some authors more optimistic about the ability of ANNs to provide understanding? One reason is that they focus on objectual understanding of a domain of things (Räz & Beisbart 2022). Such understanding consists in grasping connections in a body of knowledge or facts (Kvanvig 2003, 192), and the probabilistic correlations that ANNs capture are sufficient for this. Second, there are also examples in which networks and machine learning algorithms appear to have learned the laws of nature, e.g. Newton's law of gravitation (Lemos et al., 2023). However, a closer analysis of related research reveals that the networks were rather special and embodied significant prior knowledge of the correct explanation, for instance, the principle that the reaction equals the action. Third and finally, trained networks can, under certain conditions, be used to identify partial causes (e.g., Knüsel & Baumberger 2020). The idea is to apply Mill's method of difference (cf. Pietsch 2021, Ch. 4.3) and to run a trained ANN with inputs in which the values of some variables have been changed. Under suitable conditions, variables that impact the outputs are causes. However, applying this method requires not only non-trivial work that goes beyond simply training a model, but also some assumptions about the causes that might be relevant.

All in all, it is fair to say that the question of whether AI provides understanding has no unique answer. It depends on whether we are talking about ANNs or computer simulations, whether our interest is in explanatory or objectual understanding, and how AI is used. On the negative side, it is true that ANNs can yield accurate predictions without giving us understanding. The good news is that AI can help us to deepen physical understanding (see Krenn et al. 2022) but only if we extract the explanatory information to which some AI tools may become sensitive.

References

Baumberger, C., Beisbart, C., & Brun, G. (2016). What is understanding? An overview of recent debates in epistemology and philosophy of science. In S. Grimm, C. Baumberger, & S. Ammon (Eds.), *Explaining understanding: New perspectives from epistemology and philosophy of science.* Routledge, 1–34.

Beisbart, C. (2025). Digging deeper with deep learning? Explanatory understanding and deep neural networks. *Euro Jnl Phil Sci* **15**, 40 (2025). https://doi.org/10.1007/s13194-025-00668-y.

Burrell, J. (2016). How the machine "thinks": Understanding opacity in machine learning algorithms. *Big Data & Society*, **3**(1), Article 2053951715622512.

De Regt, H. W. (2017). *Understanding scientific understanding.* Oxford University Press.

Elgin, C. Z. (2017). *True enough*. MIT Press.

Hills, A. (2016). Understanding why. *Noûs*, **50**, 661–688. https://doi.org/10.1111/nous.12092.

Kaznatcheev, A. (2013). *Machine learning and prediction without understanding*. Theory, Evolution, and Games Group. Retrieved September 30, 2025 from https://egtheory.wordpress.com/2013/06/05/prediction-vs-understanding/.

Knüsel, B., & Baumberger, C. (2020). Understanding climate phenomena with data-driven models. *Studies in History and Philosophy of Science Part A*, **84**(C), 46–56.

Krenn, M., Pollice, R., Guo, S. Y., et al. (2022). On scientific understanding with artificial intelligence. *Nature Reviews Physics*, **4**, 761–769. https://doi.org/10.1038/s42254-022-00518-3.

Kvanvig, J. (2003). *The value of knowledge and the pursuit of understanding*. Cambridge University Press.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., et al. (2023). Learning skillful medium-range global weather forecasting. *Science* **382**(6677), 1416–1421.

Lemos, P., Jeffrey, N., Cranmer, M., Ho, S., & Battaglia, P. (2023). Rediscovering orbital mechanics with machine learning. *Machine Learning: Science and Technology*, **4**(4), Article 045002.

Pietsch, W. (2021). *Big data*. Cambridge. https://doi.org/10.1017/9781108588676.

Räz, T., Beisbart, C. (2024). The Importance of Understanding Deep Learning. *Erkenntnis* **89**, 1823–1840. https://doi.org/10.1007/s10670-022-00605-y.

Trout, J. D. (2002), Scientific explanation and the sense of understanding. *Philosophy of Science*, **69**(2), 212–233. https://doi.org/10.1086/341050.

Wittgenstein, L. (1953). *Philosophical Investigations.* Blackwell. Here following the edition: Wittgenstein, L., *Tractatus logico-philosophicus. Tagebücher 1914-1916. Philosophische Untersuchungen*, Frankfurt am Main: Suhrkamp, 1997.

Woodward, J. & Ross, L. (2021). Scientific Explanation. In: Zalta, E. N. (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2021 Edition), https://plato.stanford.edu/archives/sum2021/entries/scientific-explanation/.