



# On scientific understanding with artificial intelligence

Mario Krenn, Robert Pollice, Si Yue Guo, Matteo Aldeghi, Alba Cervera-Lierta, Pascal Friederich, Gabriel dos Passos Gomes, Florian Häse, Adrian Jinich, AkshatKumar Nigam, Zhenpeng Yao  and Alán Aspuru-Guzik 

**Abstract** | An oracle that correctly predicts the outcome of every particle physics experiment, the products of every possible chemical reaction or the function of every protein would revolutionize science and technology. However, scientists would not be entirely satisfied because they would want to comprehend how the oracle made these predictions. This is scientific understanding, one of the main aims of science. With the increase in the available computational power and advances in artificial intelligence, a natural question arises: how can advanced computational systems, and specifically artificial intelligence, contribute to new scientific understanding or gain it autonomously? Trying to answer this question, we adopted a definition of ‘scientific understanding’ from the philosophy of science that enabled us to overview the scattered literature on the topic and, combined with dozens of anecdotes from scientists, map out three dimensions of computer-assisted scientific understanding. For each dimension, we review the existing state of the art and discuss future developments. We hope that this Perspective will inspire and focus research directions in this multidisciplinary emerging field.

Artificial intelligence (AI) has been called a revolutionary tool for science<sup>1,2</sup> and it has been predicted to play a creative role in research in the future<sup>3</sup>. In the context of theoretical chemistry, for example, it is believed that AI can help solve problems “in a way such that the human cannot distinguish between this [AI] and communicating with a human expert”<sup>4</sup>. However, this excitement has not been shared by all scientists. Some have questioned whether advanced computational approaches can go beyond ‘numerics’<sup>5–9</sup> and contribute on a fundamental level to gaining of new scientific understanding<sup>10–12</sup>.

In this Perspective, we discuss how advanced computational systems, and AI in particular, can contribute to scientific understanding: we overview what is currently possible and what might lie ahead. In addition to the review of the literature, we surveyed dozens of scientists working at the interface of biology, chemistry or physics on the one hand, and AI and advanced

computational methods on the other. These personal narratives (see Supplementary Information) focus on the concrete discovery process of ideas and are a vital augmentation of the scientific literature. We discuss the literature overview and personal accounts in the context of the philosophical theory of scientific understanding recently developed by Dennis Dieks and Henk de Regt<sup>12,13</sup>. We then identify three fundamental dimensions for AI contributing to new scientific understanding (FIG. 1). (We encapsulate all advanced artificial computational systems under the term AI, independent of their working principles. In this way, we are focusing on the operational objective rather than the methodology.) First, AI can act as an instrument revealing properties of a physical system that are otherwise difficult or even impossible to probe. Humans then lift these insights to scientific understanding. Second, AI can act as a source of inspiration for new concepts and ideas that are subsequently

understood and generalized by human scientists. Third, AI acts as an agent of understanding. AI reaches new scientific insight and — importantly — can transfer it to human researchers. Although there have not yet been any examples of AI acting as a true ‘agent of understanding’ in science, we outline important characteristics of such a system and discuss possible ways to achieve it.

In the first two dimensions, the AI enables humans to gain new scientific understanding, whereas in the last, the machine gains understanding itself. Distinguishing between these classes allows us to map out a vibrant and mostly unexplored field of research, and will hopefully guide direction for future AI developments in the natural sciences.

The focus of this Perspective is how advanced computational systems and AI specifically can contribute to new scientific understanding. There are many related, interesting topics that we cannot cover here. For example, we will not discuss the relationship between scientific understanding and cognitive science, but refer the reader to a good overview<sup>14</sup>. Furthermore, we will only discuss ‘understanding’ in the context of the natural sciences, in which we can use concrete criteria from the philosophy of science and, therefore, will not touch on ‘understanding’ in a broader context (such as understanding by babies and animals, language understanding in AI and related topics). Many other works contribute to related questions and should be mentioned here. One important field of research in AI is explainable AI, which aims to interpret and explain how advanced AI algorithms come up with their solutions; see, for instance, REFS.<sup>15–18</sup>. Whereas it is not necessary, and we believe also not sufficient, to interpret the internal workings of the AI to get new scientific understanding, many of these tools and techniques can be very useful. We will briefly explain them below with concrete examples in the natural sciences. AI pioneer Donald Michie classified machine learning (ML) into three classes: weak, strong and ultrastrong, in which ultrastrong requires the machine to teach the human<sup>19</sup>. The ultrastrong ML is related to the idea of agent of understanding,